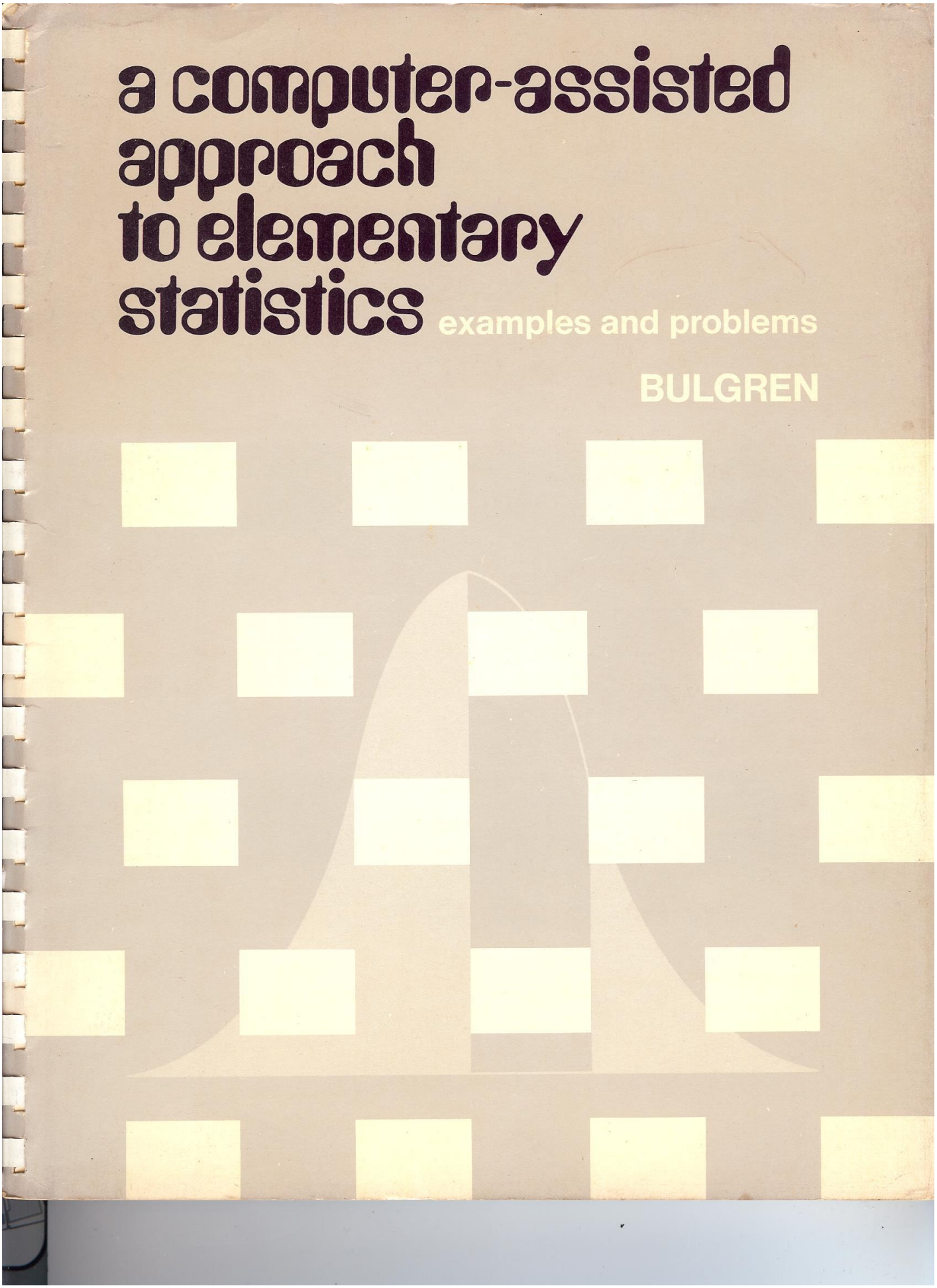
a computer-assisted approach to elementary statistics

examples and problems

BULGREN



http://www.brainless.org/Altair/

550 6F
University of Minn.
Bookstore
96

http://www.brainless.org/Altair/

A Computer-Assisted Approach to Elementary Statistics
Examples and Problems

http://www.brainless.org/Altair/

A Computer-Assisted Approach to Elementary Statistics
Examples and Problems

William Bulgren

University of Kansas

Wadsworth Publishing Company, Inc.

Belmont, California

http://www.brainless.org/Altair/

© 1971 by Wadsworth Publishing Company, Inc., Belmont, California 94002. All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transcribed, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

L. C. Cat. Card No.: 78-137147

Printed in the United States of America

1 2 3 4 5 6 7 8 9 10—75 74 73 72 71

http://www.brainless.org/Altair/

Preface

This book is designed for classroom and laboratory use in the study of elementary statistics. It provides a logical, graded, tested series of computer-oriented exercises that are intended to encourage thinking in terms of the meaning of statistics, the statistical methods, and the results of statistical analysis.

A belief that the best learning situations involve the student personally and subjectively in the subject matter has guided the preparation of this book. Programmed texts, computer-assisted learning, and other learning media are now used to accomplish this involvement. Such techniques demand that students do more than passively observe a series of lectures on their subject matter. Introductory courses in statistics lend themselves particularly well to laboratory work. Surprisingly, however, those offered by most colleges today are organized on a lecture basis—to the dissatisfaction, I believe, of student and teacher alike. Students become dissatisfied with traditional approaches to elementary statistics, because the important concepts of statistics are not easily learned by hearing someone talk about them or by reading about them in a textbook. One learns these things best, it seems, by doing the work.

The instructor will find this book so arranged that it may be used with most elementary statistics textbooks in either a one-semester or a full-year course.

Preface

http

I appreciate the helpful suggestions and criticism offered by many students, especially Kenneth Norland of the University of Kansas; and I am indebted to my wife, Janis, for her help in preparing the manuscript and to Gayle Johnson for typing it. I am also indebted to Earl J. Schweppe, Chairman of the Computer Science Department, and to Paul Wolfe, Director of the Computation Center, University of Kansas. Lastly, Alexander Kugushev, Mathematical Sciences Editor of Wadsworth Publishing Company deserves the credit for getting me started on such an enterprise as this book in the first place. I have received from many and return this book to those who find an interest or need.

http

Contents

1

Introduction 1

General discussion of statistics and computers; explanation of terminology; explanation of appendix; remarks on the exercises.

2

Describing Distributions of Measurements 5

Basic methods for describing sets of measurements.

Contents

http

3

Probability—Basic Concepts 15

Introduction to probability concepts, laws of probability, and counting techniques.

4

Random Variables and Probability Distributions 23

Introduction to random variables; mathematical expectation; hypergeometric distribution; binomial distribution; Poisson distribution; uniform distribution; normal distribution.

5

Statistical Inference—Large Sample Methods 37

Large sample methods for estimating and testing hypotheses concerning population means and the difference between two means.

6

Statistical Inference—Exact Distributions 49

The chi-square, t and F distributions; methods for estimating and testing hypotheses concerning population means, the difference between two means, a population variance, and a comparison of two population variances.

7

Linear Regression 61

Simple linear regression; statistical inferences in regression analysis.

http

8

Analysis of Variance 67

Analysis of variance; statistical inferences in analysis of variance.

Appendix I

Basic Elements of FORTRAN IV 73

Appendix II

Generating Pseudorandom Numbers, U(0, 1) 85

Appendix III

Generating Normal Random Numbers 91

Appendix IV

Answers to Selected Examples and Exercises 95

http

1

Introduction

The use of the computer in education has stimulated a great deal of interest and perhaps some unjustified anxiety concerning the computer's effect on education. This book outlines a set of exercises for use in a noncalculus elementary statistics course. For those who still have doubts about the ultimate use of computers in statistical education or who fear the effect of these techniques on the teaching of statistics, it is hoped that the following chapters will show the potential savings in time, increase in efficiency, and a new type of freedom which these new tools offer.

The book consists of eight chapters plus four appendices. Each chapter discusses important concepts and principles of statistics. Each also includes a discussion of the prerequisites for the chapter, a brief introduction with one or more examples for illustration, and a list of exercises to provide greater insight and understanding of the process being illustrated.

There are ample exercises for courses of one to four semester hours. Although the student probably will be unable to write more than two or three programs from each chapter in the allotted time, he will learn a great deal by reading and thinking about the Statement of the Problem and the Discussion and Questions of each exercise.

Four appendices are included to aid you, the student, in successfully completing this book. To avoid objections arising from the use of a particular programming language

http

in the text, Appendix I and Appendix IV are included. Appendix I gives the essential parts of FORTRAN programming. Your competence will increase as you continue to practice this skill. Although much of FORTRAN is not explained in Appendix I, you will be able to resolve the remaining mysteries by reading any complete FORTRAN manual. The following is a list of some *typical* FORTRAN programming books.

1. Anderson, Decima M., *Computer Programming: FORTRAN IV.* Appleton-Century-Crofts, 1966.
2. Dimitry, Donald L., and Mott, Thomas H. Jr., *Introduction to FORTRAN IV Programming.* Holt, Rinehart and Winston, 1966.
3. McCracken, Daniel D., *A Guide to FORTRAN IV Programming.* John Wiley & Sons, Inc., 1965.
4. Organick, Elliot I., *FORTRAN IV Primer.* Addison-Wesley Publishing Co., Inc., 1966.

It is also suggested that manufacturers' manuals on FORTRAN be consulted, along with any supplementary material pertaining to the particular installation you plan to use, especially the program library.

Appendix IV includes programs written by students that give answers to selected examples and exercises in FORTRAN. In practice, you will find that once a program has been written, you will modify and expand it several times to account for various alternatives. Many of the exercises add to exercises that have already been completed earlier. This program growth exists from exercise to exercise within each chapter, and occasionally from chapter to chapter; hence, all programs should be saved until you have finished the book.

Appendix II is concerned with generating pseudorandom numbers. These numbers represent experimental data with which you will perform computer experiments as described in Chapters 2-8. The appendix summarizes a technique to generate pseudorandom numbers and presents a FORTRAN function as well as a table consisting of 1000 random digits.

Appendix III is concerned with generating normal random numbers. These numbers represent experimental data with which you will perform computer experiments as described in Chapters 2-8. The appendix summarizes a technique to generate normal random numbers and presents a FORTRAN function as well as a table consisting of 1000 normally distributed random numbers.

Since this book will be used as a supplement to an introductory textbook in elementary statistics, the following table illustrates which chapters of this book correspond to chapters of some typical elementary statistics textbooks. Since all elementary textbooks contain the basic elements included in this book, the table is only a guideline and by no means lists all the excellent textbooks available on the market.

http

Bulgren	[1] Alder & Roessler	[2] Freund	[3] Hoel	[4] Huntsberger	[5] Mendenhall
2	2, 4	2, 3, 4	2	2, 3	3
3	5	5	3	4	4
4	6, 7, 8	6, 7, 8	4, 5	5, 6	5, 6, 7
5	9	9, 10	6, 7	6, 7, 8, 9	8
6	10, 13, 16	9, 10	6, 7	7, 8, 9	9
7	12	14	9	11	10
8	17	12	12	12	13

For example, Chapter 2 of this book (Bulgren) would be utilized with Chapter 3 of Mendenhall [5], Chapters 2 and 4 of Alder and Roessler [1], and so forth.

Typical Elementary Textbooks

1. Alder, Henry L., and Roessler, Edward B., *Introduction to Probability and Statistics,* 4th ed. W. H. Freeman and Co., 1968.
2. Freund, John E., *Modern Elementary Statistics*, 3rd ed. Prentice-Hall, Inc., 1967.
3. Hoel, Paul G., *Elementary Statistics*, 2nd ed. John Wiley & Sons, Inc., 1966.
4. Huntsberger, David V., *Elements of Statistical Inference.* Allyn and Bacon, Inc., 1967.
5. Mendenhall, William, *Introduction to Probability and Statistics,* 3rd ed. Wadsworth Publishing Co., Inc., 1971.

http

2

Describing Distributions of Measurements

Prerequisites

The exercises of this chapter require a basic understanding of methods for describing distributions of measurements. Since the computer is also being utilized, you will be studying a technique which forces you to organize your thoughts concerning a particular problem and to carefully plan a method for solving a problem before writing the actual computer program for solution.

Introduction

You have already seen the two basic methods for describing sets of measurements: graphical and numerical. Examples of graphical methods are frequency histograms and relative frequency histograms. Numerical descriptive methods involve such things as measures of central tendency (for instance, the mean) and measures of variability (for instance, the variance).

Example 2.1. The customary way to organize data derived from observations is to display them as a frequency histogram, which shows the number of times the variable falls in different intervals or within certain limits. Table 2-1 illustrates a frequency table of pseudorandom numbers contained in the interval from zero (0)

http

Table 2-1 Frequency Table of Pseudorandom Numbers

i	Interval	Midpoint (X)	Frequency (f_i)	Relative Frequency (f_i/n)	Cumulative Relative Frequency
1	0.0-0.1	.05	5	0.10	0.10
2	0.1-0.2	.15	2	0.04	0.14
3	0.2-0.3	.25	6	0.12	0.26
4	0.3-0.4	.35	7	0.14	0.40
5	0.4-0.5	.45	2	0.04	0.44
6	0.5-0.6	.55	5	0.10	0.54
7	0.6-0.7	.65	4	0.08	0.62
8	0.7-0.8	.75	9	0.18	0.80
9	0.8-0.9	.85	6	0.12	0.92
10	0.9-1.0	.95	4	0.08	1.00
			50		

to one (1), which is denoted by U(0,1). By adding all the frequencies contained in Table 2-1, one finds the total number of observations to be 50. Now the end points of the classes (0.0, 0.1, 0.2, 0.3, and so forth, in Table 2-1) are called *class limits* or *class boundaries.* The midpoint between the upper and lower class boundaries is called the *class mark* or *class midpoint.* The difference between the largest and smallest class boundaries is called the range of the table. (In Table 2-1, the range is 1.0-0.0 = 1.0.) The number of observations (variates) falling within a given class interval is called the *class frequency,* or simply, *frequency.* (In Table 2-1, the frequency for class 1 is 5, or $f_1 = 5$.) The set of frequencies with their respective classes or class midpoints is called a *frequency distribution* (Table 2-1 is such a distribution.) Experience shows that in classifying data in a frequency table, a number between 10 and 20 classes is desirable. When the number of classes is less than 10, much accuracy is lost, and when the number of classes is greater than 20, the obtained summary is too extensive.

The number of observations (variates) falling within a given class interval divided by the total number of observations is called the *relative class frequency* or *relative frequency.* (In column 5 of Table 2-1, the relative frequency for class 1 is .10, or $f_1/50 = 5/50 = .10$.) The set of relative frequencies with their respective classes or class midpoints is called a *relative frequency distribution.* The *cumulative frequency (cumulative relative frequency)* for a given class is the sum of frequencies (relative frequencies) of all classes up to and including the given class, assuming that the class with the smallest class limits is written first. Table 2-1 illustrates only the cumulative relative frequency.

Histograms for the data of Table 2-1 are shown in Figures 2-1, 2-2, and 2-3. In Figure 2-1, called a *frequency histogram,* the frequency corresponding to a class is represented by the height of a rectangle whose base is the class interval. In Figure 2-2, called a *relative frequency histogram,* the relative frequency corresponding to a class is represented by the height of a rectangle whose base is the class interval. In Figure 2-3, called a *cumulative relative frequency histogram,* the

http

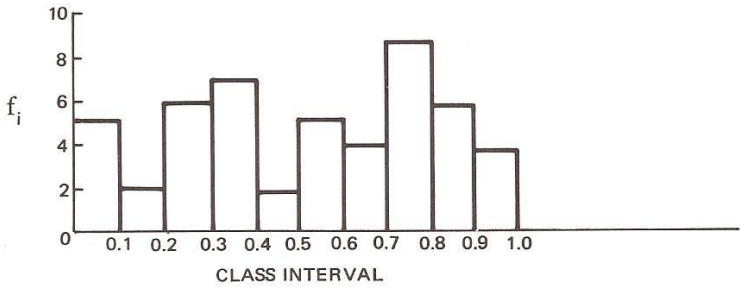


Figure 2-1 Frequency Histogram

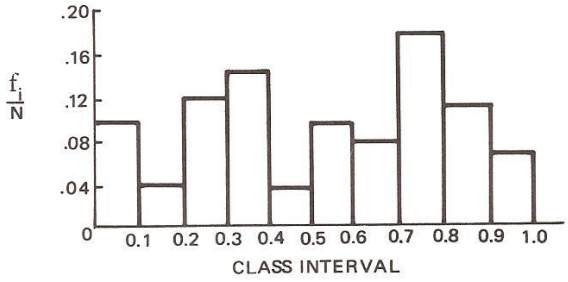


Figure 2-2 Relative Frequency Histogram

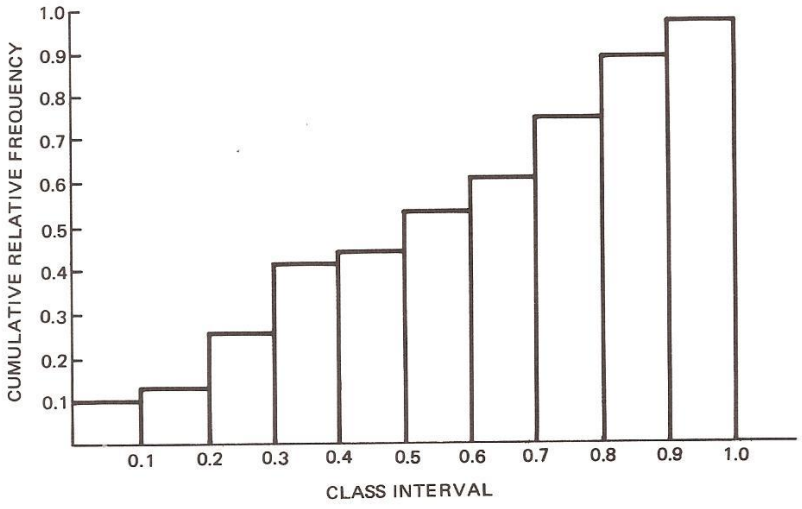


Figure 2-3 Cumulative Relative Frequency Histogram

cumulative relative frequency corresponding to a class is represented by the height of a rectangle whose base is the class interval.

Another way to represent a set of observations is by using a frequency polygon. A broken-line graph having class midpoints represented on the horizontal axis and frequencies on the vertical axis is called a *frequency polygon.* A frequency polygon for the data of Table 2-1 is shown in Figure 2-4.

If the number of class intervals increases, while at the same time the total number of observations is increased, the histogram may approach a curve, called a *frequency curve*. This curve is approximated by fitting a smooth curve to the points of a frequency polygon.

Next consider a scheme for representing any frequency distribution. Such a generalized scheme must be capable of representing a frequency distribution with any number of intervals and involving any number of values. The score value corresponding to the midpoint of a given interval will be represented by X, and the identification number for that class is affixed as a subscript, that is, X_i. It should be clear that

$$n = \sum_{i=1}^{c} f_i = f_1 + f_2 + \cdots + f_c ,$$

where c denotes the number of classes. In our example, $50 = 5 + 2 + \cdots + 4$. Note that

$$\sum_{i=1}^{c} \frac{f_i}{n} = 1.$$

In terms of the scheme under consideration, the sum of the f_1 scores in interval 1 is $f_1 X_1$, the sum of the f_2 scores in interval 2 is $f_2 X_2$, and so forth. Thus, the sum of

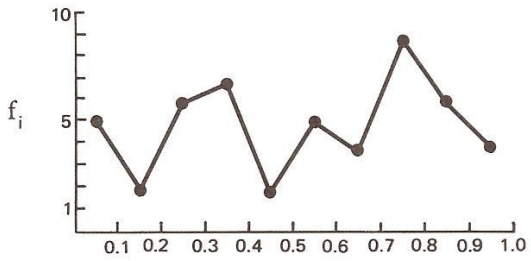


Figure 2-4 Frequency Polygon

http

the n scores involved in any frequency distribution may be represented by

$$\sum_{j=1}^{c} f_j X_j = f_1 X_1 + f_2 X_2 + \cdots + f_c X_c.$$

Similarly, the sum of squares of the n scores of any frequency distribution is represented by

$$\sum_{j=1}^{c} f_j X_j^2 = f_1 X_1^2 + f_2 X_2^2 + \cdots + f_c X_c^2.$$

From Table 2-1,

$$\sum_{j=1}^{10} f_j X_j = 5 \times .05 + 2 \times .15 + \cdots + 4 \times .95 = 26.4$$

and

$$\sum_{j=1}^{10} f_j X_j^2 = 5 \times (.05)^2 + 2 \times (.15)^2 + \cdots + 4 \times (.95)^2 = 17.905.$$

The approximate character of the equivalence of these sums to the corresponding sum derived from the ungrouped scores,

$$\sum_{i=1}^{n} X_i, \quad \sum_{i=1}^{n} X_i^2,$$

should be clearly recognized. The limits of summation have been retained here to point out that the approximately equivalent sums do not involve the same number of terms. In our example,

$$\sum_{i=1}^{50} X_i \; 26.518 \approx \sum_{j=1}^{10} f_j X_j = 26.4$$

and

$$\sum_{i=1}^{50} X_i^2 = 18.18 \approx \sum_{j=1}^{10} f_j X_j = 17.905.$$

http

It must be pointed out that as the number of intervals increases and at the same time, the total number of observations is increased, this approximation will become better.

Exercise 2.1

Statement of the Problem. Prepare and execute a program that will take n observations from a population which is U(0,1). Make a frequency table, with c intervals, showing frequency, relative frequency, relative cumulative frequency, and histogram. (It is not necessary to perform the construction of a histogram on the computer.) Find the mean and variance.

Input. Test your program with the following sets of values. Use any format you consider convenient.

Data Set 1	*Data Set 2*	*Data Set 3*
n = 100	n = 200	n = 200
c = 10	c = 10	c = 20

Output. Your program should output a frequency table showing frequency, relative frequency, and relative cumulative frequency. Also output the sum, sum of squares, mean, variance, and standard deviation for each data set.

Discussion and Questions. Since we are taking observations from a U(0,1), there should be an equal number of observations in each interval. Hence there should be n/c observations in each interval. For example, in Data Set 1, n/c = 100/10 = 10. If indeed we are observing values from U(0,1), then each interval should be approximately n/c values.

1. Why will we not obtain exactly n/c values in each interval?
2. Why should the number of values in each interval be closer to n/c in Data Set 2 than in Data Set 1?
3. How do the grouped and ungrouped mean and variance compare for Data Sets 2 and 3? (For a U(0,1), the mean is equal to 1/2 and the variance 1/12).
4. How do the mean and variance compare for Data Sets 1 and 2?
5. If the frequency in each interval is n/c, what is the relative frequency in each interval?
6. If n were equal to 1000, would you change your program? If yes, how and why?

Exercise 2.2

Statement of the Problem. Prepare and execute a program that will take n observations from a population which is N(5,10). From these observations, find the mean and variance. Subtract a constant k_1 from each of the given observations, and find the mean and variance of the new observations. Multiply each of the original observations by a constant k_2 and find the mean and variance of the new observations.

http

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
$n = 100$	$n = 100$
$k_1 = 1.0$	$k_1 = 5$
$k_2 = 5.0$	$k_2 = 1/\sqrt{10}$

Output. Your program should output the mean and variance of the original observations. Also output each mean and variance obtained from the new observations.

Discussion and Questions. Since we are taking observations from a population which is N(5,10), the mean and variance of the population are 5 and 10, respectively. Since we are considering n = 100 observations, the calculated value of the mean and variance should be approximately 5 and 10, respectively.

1. What are the effects on the mean and variance of subtracting and/or multiplying a constant from the original observations?
2. What is the significance of using Data Set 2 to test your program?
3. What in your opinion would be the effect on the mean and variance if the number of observations were increased to n = 1000?

Exercise 2.3

Statement of the Problem. Prepare and execute a program to perform a coin-tossing experiment on the computer. Let u denote a single observation from U(0,1), $0 < u < 1$. If $u \leq .5$, then say the coin is a tail or a 0. If $u > .5$, then say the coin is a head or a 1. Using this process, write a program that will toss ten coins and record X, the number of heads observed. Repeat this process n times, thus providing n values of X. Construct a relative frequency histogram for these measurements. (It is not necessary to perform this construction on the computer.)

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
$n = 50$	$n = 100$

Output. Your program should output the number of heads observed for each of the n repetitions of the experiment.

Discussion and Questions. As would be expected, the theoretical relative frequencies for this problem are known. Table 2-2 illustrates the theoretical relative frequency (called probability) associated with this coin-tossing experiment.

1. Compare your observed relative frequencies to Table 2-2.
2. What is the difference between Data Sets 1 and 2? Does Data Set 2 give answers closer to the theoretical relative frequencies? Why?

http

Table 2-2

X	Theoretical Probability
0	1/1024
1	10/1024
2	45/1024
3	120/1024
4	210/1024
5	252/1024
6	210/1024
7	120/1024
8	45/1024
9	10/1024
10	1/1024 1 = SUM

3. If ten coins are repeatedly tossed a large number of times, then on the average you would expect to see approximately 5 heads. What is the mean for this experiment when one uses either n = 50 or n = 100?
4. If you were going to bet a large amount of money on the number of heads showing in one toss of ten coins, what value of X would you choose? Why?

As a further exercise, construct and graph the cumulative relative frequencies for this exercise.

Exercise 2.4

Statement of the Problem. Prepare and execute a program that will take n observations from a population which is N(0,1). From these observations, find the fraction of observations falling in the intervals $\mu \pm \sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$. ($\mu = 0$, $\sigma^2 = 1$), and make a relative frequency histogram for the six intervals, namely,

$$\mu - 3\sigma \text{ to } \mu - 2\sigma, \quad \mu - 2\sigma \text{ to } \mu - \sigma, \quad \mu - \sigma \text{ to } \mu, \quad \mu \text{ to } \mu + \sigma, \text{ etc.}$$

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
n = 100	n = 200

Output. Your program should output the number of observations in each of the six intervals.

Discussion and Questions. Tchebysheff's (Chebyshev's) Theorem states that for $k \geqslant 1$ and n observations, at least $(1 - 1/k^2)$ of the observations are in the interval $\mu \pm k\sigma$. For example, with k = 2, at least 3/4 of the observations are in the interval $\mu \pm 2\sigma$. The Empirical Rule states that approximately 68% of the observations are in the interval $\mu \pm \sigma$, approximately 95% of the observations are in the interval $\mu \pm 2\sigma$, and approximately 99.7% of the observations are in the interval $\mu \pm 3\sigma$.

http

1. Compare your results with Tchebysheff's Theorem.
2. Compare your results with the Empirical Rule.
3. Which of the above two rules gives a better estimate of your results? Why?
4. Why does Tchebysheff's Theorem underestimate your results?
5. Construct relative frequency histograms for your results, for Tchebysheff's Theorem, and for the Empirical Rule.

Exercise 2.5

Statement of the Problem. Prepare and execute a program which will use the process described in Exercise 2.3. Using this process, write a program that will toss ten coins and record X, the number of heads observed. Repeat this process n times. Calculate the accumulated proportion of heads after each ten coins and plot your results.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*	*Data Set 3*	*Data Set 4*
n = 50	n = 100	n = 150	n = 200

Output. Your program should output the relative frequency of heads after each group of ten coins has been flipped.

Discussion and Questions. The purpose of this exercise is to show that the accumulated proportion of heads fluctuates, but it comes closer and closer to a value of 1/2.

1. Does increasing n from 100 to 150 affect the accumulated results? How?
2. Does increasing n from 50 to 150 affect the accumulated results? How?
3. Which data sets would you expect to illustrate the purpose more clearly? Why?

http

3

Probability

Prerequisites

The exercises of this chapter require a basic understanding of probability. This subject includes the concepts of a sample space and events, the Additive Law of Probability, complementary events and partitions, conditional probability, Law of Multiplication, independent events, and counting techniques.

Introduction

Frequently we see statistics divided into two parts, descriptive statistics and inductive statistics. As we saw from the previous chapter, descriptive statistics is concerned with describing a set of measurements through such methods as graphic and tabular presentation, and calculating numerical measures such as the mean and variance. Statistical inference, or inductive statistics, is concerned primarily with making generalizations and decisions about a population based on sample information.

The definition of statistics to be used here is that statistics is the science of inductive inference. Clearly this definition focuses attention on induction, since it is reasonable to assume that the primary purpose of gathering data is to

http

generalize. This definition is not intended to downgrade the importance of descriptive statistics.

Whenever we are faced with the problem of making decisions from incomplete information, we face the risk of making incorrect decisions. Since we wish to make the best possible decision based on the information available, our statements concerning the true state of affairs must be put forward in probabilistic terms. Therefore, in order to understand inductive statistics, it is essential to understand probability.

Example 3.1. In order to demonstrate the role probability plays in modern decision making processes, let us begin with a game of chance. For example, consider a coin-tossing experiment on the computer. Let u denote a single observation from U(0,1), $0 < u < 1$. If $u > .5$, then say the coin is a head or a 1. If $u \leqslant .5$, then say the coin is a tail or a 0. Using this process, we write a program that will toss two coins and record X, the number of heads observed. We have just formed a model of an experiment in the sense that every possible outcome of the experiment can be represented by an element or a point. Figure 3-1 illustrates the model of our experiment. The term *sample space* will be used synonymously with the term "model of an experiment." Hence every possible outcome of our experiment can be represented by an element or point in the sample space.

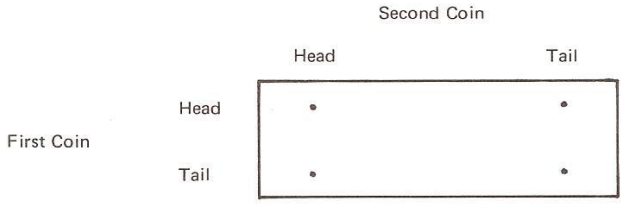


Figure 3-1

In our example, the sample space has four equally likely and mutually exclusive points (outcomes). Hence each of the four points has a probability of 1/4 associated with it. Let us now carry out our experiment 100 times. Table 3-1 gives the results.

Table 3-1

		Second coin: H	Second coin: T
First coin	H	23	26
	T	24	27

http

It should be clear from Table 3-1 that our computer experiment yields a value of 23/100 = .23 for an estimate of the probability of exactly two heads. Why is this probability not equal to 1/4 = .25? In our situation, *a priori* considerations lead to a complete specification of the probability model—that is, the probability of selecting a particular sample point in the sample space was 1/4. In our computer experiment, we thought of probability as being the relative frequency with which a particular outcome occurs. Hence, the value .23 was based on *a posteriori* considerations. Situations do exist where intuition fails and the complete specification of the model on *a priori* grounds is not feasible. As an example of such a situation, consider cutting a globe of the earth into two parts. What is the probability that if one of the parts is tossed into the air, the half-sphere will come to rest with the flat side up?

Exercise 3.1

Statement of the Problem. Prepare and execute a program to perform an experiment which rolls a pair of dice. Let u_1 denote a single observation from U(0,1), $0 < u_1 < 1$. If $0 < u_1 < 1/6$, then the number showing upright on the first die is a 1. If $1/6 \leq u_1 < 2/6$, then the number on the first die is a 2. If $2/6 \leq u_1 < 3/6$, then the number on the first die is a 3. If $3/6 \leq u_1 < 4/6$, then the number on the first die is a 4. If $4/6 \leq u_1 < 5/6$, then the number on the first die is a 5. If $5/6 \leq u_1 < 1$, then the number on the first die is a 6. Let u_2 denote another single observation from U(0,1). Generate the outcome of the throw of the second die exactly like the first die. Repeat the process of rolling a pair of dice n times. Count the number of occurrences of each possible outcome.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
n = 200	n = 500

Output. Your program should output the number of occurrences of each possible outcome.

Discussion and Questions. As would be expected, *a priori* considerations lead to a complete specification of the probability model. Since we have a balanced six-sided die numbered in the usual manner and it is to be rolled twice, our sample space consists of 36 points. An appropriate probability model is a set of 36 sample points, with a probability of 1/36 associated with each.

1. Compare the number of occurrences of each possible outcome with the numbers to be expected on theoretical grounds, using both data sets.
2. What is the difference between Data Sets 1 and 2? Does Data Set 2 give answers closer to the theoretical answers? Why?
3. What is your observed value for the probability of obtaining a one on each of the die ("snake eyes")? How does it compare with the theoretical result?
4. What is your estimate of the probability of obtaining a seven? How does it compare with the theoretical result?

http

5. Construct a relative cumulative frequency histogram for the number of points on the two dice. How does it compare with the theoretical results?

Exercise 3.2

Statement of the Problem. Prepare and execute a program to perform an experiment which rolls a die and tosses a coin. The experiment will be performed as a combination of Example 3.1 and Exercise 3.1 except that only one die and one coin will be used. Repeat the experiment, "rolling a die and tossing a coin" n times and count the number of occurrences of each possible outcome.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
n = 100	n = 500

Output. Your program should output the number of occurrences of each possible outcome.

Discussion and Questions. Since we have a balanced six-sided die numbered in the usual manner and an unbiased two-sided coin, our sample space consists of 12 points with a probability of 1/12 associated with each.

1. Compare the number of occurrences of each possible outcome with the numbers to be expected on theoretical grounds.
2. What is your observed value for the probability of a "head on the coin and a six on the die"? How does it compare with the theoretical result?
3. Does Data Set 2 yield answers closer to the theoretical answers? Why?
4. Does your observed value for the probability of a "tail on the coin and a six on the die" yield the same answer as Question 2? Why?

Exercise 3.3

Statement of the Problem. As in Example 3.1, construct a coin tossing experiment on the computer. Using this process prepare and execute a program that will toss two coins and then record which one of the following events occurs:

	Coin 1	Coin 2
E_1:	Head	Head
E_2:	Head	Tail
E_3:	Tail	Head
E_4:	Tail	Tail

Repeat this process n times.

http

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
n = 100	n = 500

Output. Your program should output the number of times coin 1 is a head, the number of times coin 2 is a head, the number of times coin 2 is a head given coin 1 is a head, and the number of times both coin 1 and coin 2 are heads (that is, E_1).

Discussion and Questions.

1. Compute the following, using your observed results (Note: P denotes probability):
   a. P (coin 1 is a head)
   b. P (coin 2 is a head)
   c. P (coin 2 is a head, given that coin 1 is a head). Compare this to Part a. What conclusion does one draw?
   d. Check computationally that part c is $P(E_1)/P$(coin 1 is a head).
2. Compare the results in Question 1 with the known theoretical results.
3. What can be said about independence in this example?

Exercise 3.4

Statement of the Problem. Repeat Exercise 3.3, except assume that coin 1 is a biased coin with the probability of a head equal to .6 instead of .5. Answer the same set of questions you did in Exercise 3.3

Exercise 3.5

Statement of the Problem. The probability of a batter striking at a ball is .8. When he strikes, the probability of his hitting is .6. If he hits the ball, the probability of its being caught is .6. Prepare and execute a program to perform an experiment using U(0,1) to compute the probability of striking, hitting, and being caught. Determine how many of the first n plays will result in a batter's ball being caught.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
n = 10	n = 100

Output. Your program should output the number of occurrences of each possible outcome.

http

Discussion and Questions.

1. Compute the following, using your observed results:
   a. P (Striking)
   b. P (Hitting)
   c. P (Being caught)
2. Compare the results in Question 1 with the known theoretical results.
3. Does Data Set 2 yield answers closer to the theoretical answers? Why?

Exercise 3.6

Statement of the Problem. Prepare and execute a program to compute

$$C_r^n = \binom{n}{r} = \frac{n!}{r!\ (n - r)!}$$

and

$$P_r^n = \frac{n!}{(n - r)!}$$

where r = 0, 1, 2, ..., n.

Input. Test your program with the following sets of input values.

Data Set 1	*Data Set 2*	*Data Set 3*
n = 5	n = 10	n = 20
r = 2	r = 2	r = 2

Output. Your program should output both C_r^n and P_r^n.

Discussion and Questions. The number of permutations of n distinct objects is n! or P_r^n. Associated with any of the n objects in the first position we can have any of (n - 1) objects in the second position, any of (n - 2) in the third position, and so forth. Therefore, the number of different arrangements is

$$n(n - 1)(n - 2) \cdots 3 \cdot 2 \cdot 1 = n!.$$

The number of permutations of n distinct objects, taken r at a time, without replacement, is

$$\frac{n!}{(n - r)!} = P_r^n.$$

This statement should be clear, since we can select any of the n objects for the first position, any of the remaining (n - 1) objects for the second position and so forth. We can continue in this manner until we read the r^{th} position, which can be filled by

http

any of the remaining n − r + 1 items. Therefore, the number of different arrangements is

$$n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

The total number of combinations of n distinct objects, taken r at a time, without replacement, is

$$\frac{n!}{r!(n-r)!} = \binom{n}{r} = C_r^n.$$

We know that each combination of r distinct objects has r! arrangements. Hence, the total number of permutations, P_r^n, must equal

$$\binom{n}{r}r!, \text{ or } \binom{n}{r} = \frac{P_r^n}{r!}$$

1. Compare the answers for the three data sets with the known theoretical results.
2. Check the equality $\binom{n}{r} = P_r^n/r!$ using the three data sets.
3. Check your program with other sets of data, such as $n = 10$, $r = 5$.
4. Is there any limitation on the size of n for your computer? Why? What is it?

Exercise 3.7

Statement of the Problem Prepare and execute a program to generate Pascal's triangle. This triangle can be generated in the following manner: ones are written on a vertical leg and on the hypotenuse of an isosceles right triangle. Each additional element filling up the triangle is the sum of the elements directly above it and the element above and to the left. The numbers in the $(n+1)^{th}$ row are then the coefficients in the binomial expansion of $(X+Y)^n$, that is, C_r^n. Show how this technique can be used to compute C_r^n.

Input. Test your program with the following set of values.

Data Set 1	*Data Set 2*
n = 10	n = 20

Output. Your program should output Pascal's triangle.

Discussion and Questions. In order to facilitate the working of this exercise, the following is a picture of Pascal's triangle for n = 4.

http

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1

Hence, $\binom{4}{0} = 1, \binom{4}{1} = 4, \binom{4}{2} = 6, \binom{4}{3} = 4, \binom{4}{4} = 1$.

1. In general, the $(n + 1)^{th}$ row of Pascal's triangle contains the values of $\binom{n}{r}$ for $r = 0, 1, 2, \ldots, n$. Use this fact to calculate the values of $\binom{6}{r}$ for $r = 0, 1\ 2, 3, 4, 5, 6$.
2. Compare your results with those of Exercise 3.5.
3. Is there any limitation on the size of n for your computer? Why? What is it?

http

4

Random Variables and Probability Distributions

Prerequisites

The exercises of this chapter require a basic understanding of random variables, discrete probability distributions such as the hypergeometric, binomial, and Poisson, and continuous probability distributions such as the uniform and normal.

Introduction

As we pointed out in the previous chapter, statistical inference is concerned primarily with making generalizations and decisions about a population based on sample information. For this reason we are interested in observations generated by an experiment. Instead of using the terminology "the observations generated by an experiment," we shall use "the variable measured in the experiment." A variable whose value is known before an experiment is performed *cannot* be random. For example, if we knew a head was going to appear in a one-toss coin experiment, then the variable, the number of heads in one flip of coin, would always be the one and obviously not random.

Our interest here is in variables that are random—that is, those whose value is not known until an experiment is performed. We shall use capital letters to denote

random variables and small letters to represent a value that the random variable may take. At this point, we shall consider only the case of discrete random variables. The notation p(x) will be used to represent the probability that the random variable X takes on the value x, and $p(x) = P(X = x)$ will be called the discrete probability distribution or function of X.

Example 4.1. In order to understand the role random variables and probability distributions play in statistics, consider the game of matching pennies. Let A and B be the two individuals who are matching pennies. If A matches B, then A wins a cent; otherwise, B wins one penny. Let X be A's "winnings" in one play. The simple events will be identified by the notation HT, where the result of A's coin appears first (a head in this case) and the result of B's coin appears second (a tail in this case). We have X(HH) = X(TT) = +1 and X(HT) = X(TH) = -1; that is, if an HH or a TT appears, A wins one penny, and if an HT or TH appears, A loses one penny. It should be clear that

$$p(+1) = P(X = 1) = 1/2 \text{ and } p(-1) = P(X = -1) = 1/2.$$

If A and B play this game N times and if n matches result, then A's average winnings will be his wins minus his losses, or

$$\frac{n}{N} - \frac{(N - n)}{N} = \frac{n - (N - n)}{N} = 1 \cdot \frac{n}{N} + (-1)\left(1 - \frac{n}{N}\right)$$

Now n/N is an approximation to the probability of a match, and (1 - n/N) is an approximation to the probability of a non-match. Hence A's average winnings will be an approximation of

$$1 \cdot P(X = +1) + (-1)\ P(X = -1).$$

Therefore, we say that A will "expect" to win

$$1 \cdot P(X = +1) + (-1)\ P(X = -1) = 0$$

(cents) each time he matches coins with B. On the average A should "expect" neither to win nor lose in a long sequence of matches. In general, we define the *expectation* (mean or average value) of any discrete random variable X as

$$E(X) = x_1 P(X = x_1) + x_2\ P(X = x_2) + \cdots$$
$$= \sum_i x_i P(X = x_i) = \sum_i x_i p(x_i),$$

where X is a discrete random variable which can assume any one of several distinct values $x_1, x_2, \ldots$, with associated probabilities $p(x_1), p(x_2), \ldots$

In order to distinguish between a theoretical, or population, distribution for a random variable and the empirical distribution for a random variable, consider performing the previous example on a computer. Let u denote a single observation

http

from U(0,1), $0 < u < 1$. If $u > .5$, then let $X = +1$ and if $u \leq .5$, then let $X = -1$. Using this process, we shall write a program that will match coins and record X, the amount of A's winnings. We have just formed a model of our experiment in the sense that every possible outcome of the experiment can be represented by an element or a point. Since there is a finite number of these points, we say that X is a discrete random variable. The observed relative frequencies or probabilities are presented in Table 4-1 for 100 repetitions of the above experiment, and the corresponding theoretical relative frequencies or probabilities are presented in Table 4-2.

Table 4-1

X	−1	1
p(x)	$\frac{49}{100}$	$\frac{51}{100}$

Table 4-2

X	−1	1
p(x)	1/2	1/2

It should be clear at this point that the theoretical, or population, distribution for a random variable is the limiting form for the empirical or observed distribution as the number of repetitions of the experiment increases.

So far, we have talked about an example in which the random variable has taken on only a finite number of possible values, and each sample point in the sample space has a probability either positive or zero. For problems in which measurements are involved—length and time, for example—it is appropriate to think of a random variable as being continuous. A random variable X which is concerned with a measurement of length or of time can take on every possible value in an interval. We shall refer to p(x) as the probability density function (p.d.f.) associated with X when X is a continuous random variable. The total area under a probability density function must be equal to 1. When dealing with a continuous random variable we shall not speak of probabilities of exact values, but rather of probabilities associated with intervals. The concept of discrete versus continuous random variables will be further illustrated in the following exercises.

Exercise 4.1

Statement of the Problem. Prepare and execute a program to perform a coin-tossing experiment on the computer. Let u denote a single observation from U(0,1), $0 < u < 1$. If $u \leq .5$, then say the coin is a tail or a 0. If $u > .5$, then say the

http

coin is a head or a 1. Using this process, write a program that will toss five coins and record X, the number of heads observed. Repeat this process n times, thus providing n values of X. Compute the sample mean and sample variance and construct a relative frequency histogram for these measurements.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
n = 50	n = 200

Output. Your program should output the number of heads for each of the n repetitions of the experiment, the sample mean, and the sample variance.

Discussion and Questions. In this computer exercise, we have considered a series of experiments which have the following properties:

(a) The results of each experiment can be classified in one of two categories: head or tail.

(b) The probability, p, of a head is the same for each experiment (p = 1/2).

(c) Each experiment is independent of all the others.

(d) The experiment is performed a fixed number of times. (In our exercise the coin is tossed 5 times).

These properties are exactly those of the binomial distribution. The binomial distribution is given by the following:

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x}, & x = 0, 1, 2, \ldots, n, \\ 0, & \text{elsewhere;} \end{cases}$$

n = sample size or number of trials;

p = probability of a success on a single trial.

In our exercise, n = 5 and p = 1/2. Also, the binomial distribution is sometimes referred to as a probability distribution pertaining to sampling with replacement.

1. Compare the observed distribution of the random variable X to the theoretical results obtained from the binomial distribution.
2. What is the difference between Data Sets 1 and 2? Does Data Set 2 give answers closer to the theoretical answers? Why or why not?
3. How does the sample mean compare with the theoretical expected value? Does Data Set 2 yield closer answers than Data Set 1? Why or why not?
4. How does the sample variance compare with the theoretical variance? Does Data Set 2 yield closer answers than Data Set 1? Why or why not?
5. What is your estimate (from the program) of the probability of at least 2 heads? How does it compare with the theoretical results? Does Data Set 2 yield closer answers than Data Set 1? Why or why not?

http

Exercise 4.2

Statement of the Problem. Prepare and execute a program to perform the following experiment on the computer. Let u denote a single observation from u(0,1), $0 < u < 1$. If $u \leq .4$, then say X is 0. If $u > .4$, then say X is 1. Using this process, write a program to repeat this experiment n times and hence construct a population of n observations of 0's and 1's. Compute the mean and variance of the population and construct a relative frequency histogram.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
n = 50	n = 100

Output. Your program should output the number of 0's and 1's in the n observations.

Discussion and Questions. The purpose of this exercise is to verify that the mean of a binomial population is equal to the relative frequency of successes (1's), and the variance is equal to the product of the relative frequencies of successes and failures. From the Discussion-and-Question section of Exercise 4.1, it should be clear that the mean is the only parameter of a binomial distribution.

1. Compute the mean and variance in the usual manner.
2. Organize the observations into a frequency table. Find the mean and variance in terms of frequencies. Are these values the same as those computed in 1?
3. Draw the relative frequency histogram for the population.

Exercise 4.3

Statement of the Problem Repeat Exercise 4.1 and add a program which will find the fraction of observations falling in the six intervals $\mu - 3\sigma$ to $\mu - 2\sigma$, $\mu - 2\sigma$ to $\mu - \sigma$, $\mu - \sigma$ to μ, and so forth. Make a relative frequency histogram for the six intervals.

Input. The input is the same as Exercise 4.1.

Output. Your program should output the number of observations in each of the six intervals.

Discussion and Questions. Tchebysheff's Theorem and the Empirical Rule were stated in the Discussion and Question section of Exercise 2.4. For a binomial distribution, np is the theoretical mean and np(1-p) is the theoretical variance, where n is the number of trials and p the probability of a success on a single trial. For our particular exercise,

http

$$\mu = np = 5(1/2) = 2.5$$

and

$$\sigma^2 = np(1 - p) = 5(1/2)(1/2) = 1.25.$$

1. What does the theoretical relative frequency histogram (probability distribution) look like for the six intervals? Compare this to the theoretical probability distribution.
2. Compare both the observed and theoretical results with Tchebysheff's Theorem.
3. Compare both the observed and theoretical results with the Empirical Rule.
4. Which of the above two rules gives a better estimate of your results? Why?
5. For all the above questions, compare Data Set 2 with Data Set 1.

Exercise 4.4

Statement of the Problem. Prepare and execute a program that will take n observations from a population which is N(0,1). From these observations, find the mean and variance.

Input. Test your program with the following sets of data.

Data Set 1	*Data Set 2*	*Data Set 3*	*Data Set 4*
n = 10	n = 100	n = 500	n = 1000

Output. Output the sample mean and variance for each of the four data sets.

Discussion and Questions. It is important to note that an extremely important continuous probability density function has been introduced, namely the normal distribution. Many of the statistical techniques to be discussed later are based on normal distribution. The formula for the density function of normal distribution is

$$\frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty.$$

At this point, we shall not discuss the area under the normal curve, but only consider the two symbols μ and σ^2. Clearly, the above formula depends on μ and σ^2. Changes in the value of μ shift the curve to the left or right. Changes in the value of σ^2 change the shape of the distribution. In practice, we are interested in estimating these important parameters. We shall use the sample mean $\overline{x}$, to estimate μ and the sample variance, S^2, to estimate σ^2. As n increases in size, that is, as our sample size increases, the estimate of μ and σ^2 should approach their theoretical values.

1. Show that $\overline{x}$ approaches $\mu = 0$ as the sample size increases. Draw a graph.
2. Show that S^2 approaches $\sigma^2 = 1$ as the sample size increases. Draw a graph.

http

3. What is the difference between a small sample and a large sample in estimating a parameter of a population distribution?

Exercise 4.5

Statement of the Problem. The probabilities that 0, 1, 2, 3, 4, or 5 new products will win customer acceptance each day are, respectively, .33, .41, .20, .05, .01, and .00. Let u denote a single observation from U(0,1), $0 < u < 1$. Prepare and execute a program, where $0 \leq u < .33$ corresponds to 0 new products, $.33 \leq u < .74$ corresponds to 1 new product, $.74 \leq u < .94$ corresponds to 2 new products, and so forth, to simulate the number of new products that will be accepted on n consecutive days. Compute the average (mean) number of new products accepted per day. Compute the variance and construct the observed probability distribution.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*	*Data Set 3*
n = 20	n = 50	n = 100

Output. Your program should output the number of new products which will win customer acceptance each day for each of the n repetitions of the experiment. Output the sample mean and variance.

Discussion and Questions. We have just given a somewhat more general method that can be used to simulate any empirical distribution, any discrete distribution, or any continuous distribution that can be approximated by a discrete distribution. We are using the method here to simulate a discrete distribution, namely the binomial distribution with n = 5 and p = .20. From the computation of the theoretical binomial probability distribution, we find that

i	x_i	$P(X = x_i) = p(x_i)$
1	0	.33
2	1	.41
3	2	.20
4	3	.05
5	4	.01
6	5	.00

In order to generate X on a computer, we generate a pseudorandom number u and set $X = x_i$ if

$$p(x_1) + \cdots + p(x_{i-1}) \leq u < p(x_1) + \cdots + p(x_i),$$

where $p(x_0) = 0$. For example, assume we selected u = .51281. Now

$$.33 \leq u \leq .33 + .41 = .74.$$

Therefore, we set $X = x_2 = 1$.

We have utilized this construction in this exercise. It should be pointed out that this method is suggested for its simplicity, and in practice it is a method to be used when no other alternative is available. The method described in Exercise 4.1 yields "better" results from the standpoint of computation speeds, ease of programming, and memory requirements.

1. Compare the observed binomial probability distribution with the theoretical binomial probability distribution with n = 5 and p = .20.
2. What differences exist among the data sets? Does Data Set 3 yield better answers than the others? Why?
3. How does the sample mean compare to the theoretical expected value? Does Data Set 3 yield closer answers than the others? Why?
4. How does the sample variance compare to the theoretical variance? Does Data Set 3 yield closer answers than Data Sets 1 and 2? Why?
5. What is the estimated probability that 4 or more new products will win customer acceptance? How does this compare to the theoretical probability?

Exercise 4.6

Statement of the Problem. The probabilities that a person who spends a morning fishing at a certain spot on the East Verde river catches 0, 1, 2, 3, . . . , or 8 trout are, respectively, 0.05, 0.15, 0.22, 0.22, 0.17, 0.10, 0.05, 0.03, and 0.01. Let u denote a single observation from U(0,1), $0 < u < 1$. Prepare and execute a program, where $0 \leqslant u < .05$ corresponds to 0 trout, $.05 \leqslant u < .20$ corresponds to 1 trout, $.20 \leqslant u < .42$ corresponds to 2 trout, and so forth, to simulate the number of fish caught at this spot by a fisherman on n consecutive days. Compute the average catch per day. Compute the variance and construct the observed probability distribution.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*	*Data Set 3*
n = 20	n = 50	n = 100

Output. Your program should output the number of trout caught each day for each of the n repetitions of the experiment. Output the sample mean and variance.

Discussion and Questions. As in Exercise 4.4, we have given a method that can be used to simulate any empirical distribution, any discrete distribution, or any continuous distribution that can be approximated by a discrete distribution. We are using the method here to simulate a discrete distribution, namely, the Poisson distribution with $\lambda = 3.0$. The Poisson distribution is given by the following equation:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

http

λ = average number of occurrences

e = base of natural logarithm = 2.718. ...

1. Compare the observed Poisson probability distribution with the theoretical Poisson probability distribution with $\lambda = 3.0$, the theoretical average number of trout caught per day.
2. What is the difference among the data sets? Does Data Set 3 yield better answers than the others? Why?
3. How does the sample mean compare to the theoretical expected value? Does Data Set 3 yield closer answers than the others? Why?
4. How does the sample variance compare to the theoretical variance? Does Data Set 3 yield closer answers than Data Sets 1 and 2? Why?
5. What is the estimated probability that a day selected at random will yield a catch of at least 1 trout? How does this compare with the theoretical probability?
6. Why did we truncate the probability distribution after X = 8? What is the probability that a day selected at random will yield a catch of more than 8 trout?

Exercise 4.7

Statement of the Problem. The binomial probability distribution is used to approximate the hypergeometric distribution by letting p = M/N. This usage is justifiable when n is small compared with N. Prepare and execute a program to show that this approximation is reasonable for various values of n, N and M. (The notation used will be explained in the Discussion and Question section of this exercise.)

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*	*Data Set 3*
M = 5	M = 5	M = 5
N = 100	N = 100	N = 100
n = 5	n = 10	n = 15

Output. Your program should output the values of the binomial and hypergeometric distributions as well as the difference between the two.

Discussion and Questions. The hypergeometric probability distribution is sometimes referred to as a probability distribution that pertains to sampling *without* replacement. The hypergeometric distribution is represented as follows:

$$p(x) = P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \begin{cases} 0 \leqslant x \leqslant n & \text{if } M \geqslant n \\ 0 \leqslant x \leqslant M & \text{if } M < n \end{cases}$$

http

n = number of items in sample,

N = number of items in population,

M = number of items in population that have a particular characteristic.

The binomial distribution used in this approximation is referred to as a probability distribution pertaining to sampling *with* replacement. As a guideline, the binomial may be used to approximate the hypergeometric when N (population size) is greater than 50 and n (sample size) is less than 10% of N.

1. Compare the answers obtained from the binomial and hypergeometric distribution for the three data sets.
2. Does the guideline discussed above appear valid?
3. What justification does one have in making this approximation?

Exercise 4.8

Statement of the Problem. The Poisson probability distribution is used to approximate the binomial distribution by letting $\lambda = np$. This usage is justifiable when n is large and p is relatively close to zero. Prepare and execute a program to show that this approximation is reasonable. (The notation used is explained in the Discussion-and-Question section of this exercise.)

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*	*Data Set 3*
p = .05	p = .05	p = .05
n = 5	n = 10	n = 15

Output. Your program should output the values of the Poisson and binomial distributions as well as the difference between the two.

Discussion and Questions. The Poisson distribution is represented as follows:

$$p(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \cdots$$

λ = average number of occurrences.

The Poisson distribution may be thought of as the limit of the binomial distribution as p approaches 0, n approaches infinity, and np remains constant. Hence, the Poisson distribution may be used to approximate the binomial distribution when n is large and p is close to 0. As a guideline, the Poisson may be used to approximate the binomial when n (number of trials) is as low as 50 and p (probability of a success on a single trial) is as large as .10. The approximation is better if n is larger than 50 and p is smaller than .10.

http

1. Compare the answers obtained from the Poisson and binomial distribution for the three data sets.
2. Does the guideline discussed above appear valid?
3. What justification does one have in making this approximation?

Exercise 4.9

Statement of the Problem. Prepare and execute a program that will take n observations from a population which is N(0,1). Make a frequency table with c intervals in the interval from -4 to +4 showing frequency, relative frequency, and relative cumulative frequency. Construct a relative frequency histogram. (It is not necessary to perform the construction of a histogram on the computer.) Compute the sample mean and variance.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
n = 100	n = 500
c = 8	c = 24

Output. Your program should output a frequency table showing frequency, relative frequency and relative cumulative frequency. Also output the sum, sum of squares, mean, variance, and standard deviation for each data set.

Discussion and Questions. We shall now once again make the distinction between two general classes of random variables, *discrete* and *continuous*. The discrete distributions, such as the hypergeometric and binomial, were concerned with sample spaces in which each sample point had a probability either positive or zero. As we saw in the Poisson distribution, it is possible for the number of sample points associated with a discrete distribution to be infinite in number. In particular, the random variable takes on as many values as there are positive integers. In contrast with discrete random variables, a continuous random variable is assumed to be able to take on any value within an interval or a set of intervals. Given a random variable X associated with a normal distribution or any continuous distribution, there is in the population a distribution for various values of X which in many respects is similar to the observed discrete probability distributions we have constructed previously. The theoretical or population distribution for a random variable is the limiting form for the empirical or observed distribution as the sample size increases without limit. The purpose of this exercise is to show that if we continue to increase the sample size and if we use smaller and smaller intervals to cover the range (-4 to +4), it seems reasonable to assume that as the sample size increases without limit and the lengths of the intervals approach zero, the relative frequency histogram should approach a continuous curve like that of Figure 4.1.

The total area under a probability density function such as the normal must be equal to 1. We will not speak of probabilities of exact values, but rather of probabilities associated with intervals.

http

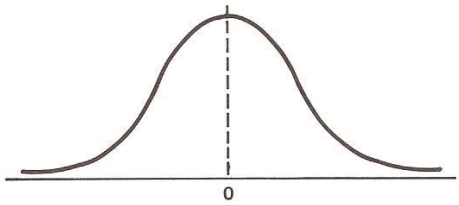


Figure 4.1 Normal distribution with mean 0 and variance 1

1. How does the observed probability distribution compare with the theoretical probability density function for each data set?
2. What is the effect on the histogram from Data Set 1 to Data Set 2?
3. Is Data Set 2 a better approximation of the theoretical distribution than Data Set 1? Why?
4. What percentage of the observations are in the six intervals $\mu - 3\sigma$ to $\mu - 2\sigma$, $\mu - 2\sigma$ to $\mu - \sigma$, $\mu - \sigma$ to μ, etc., and how does this percentage compare to the Empirical Rule?
5. How do the mean and variance compare for Data Sets 1 and 2? How do these values compare to the theoretical values of $\mu = 0$ and $\sigma^2 = 1$?
6. In dealing with continuous random variables, why do we say that the probability of an exact value is zero?

Exercise 4.10

Statement of the Problem. Prepare and execute a program that will take M sets of random samples, each consisting of n observations, from a population which is N(5,100). Compute the sample mean and variance for the n times M observations as well as the sample mean and variance for the M sample means, each consisting of n observations. Construct a histogram for the distribution.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
M = 200	M = 200
n = 5	n = 10

Output. Your program should output a frequency table for the sample means, the mean and variance of the total sample, and the mean and variance of the sample means.

Discussion and Questions. One of the most important theorems concerning sampling distributions is the *Central Limit Theorem*. The theorem states that if random samples are selected from a population with mean μ and finite variance σ^2, then as the sample size increases without limit, the distribution of $\overline{X}$ (the sample mean) approaches a normal distribution with mean μ and variance σ^2/n.

http

Another less formal manner of stating the above is that means of samples of random measurements drawn from a population tend to possess, approximately, a bell-shaped distribution in repeated sampling. It is worth stating again that the original population does not need to be normal as is the case in this exercise.

1. How many intervals are needed to illustrate the observed probability distribution?
2. Draw the histogram for the sampling distribution and compare the shape with that of the population.
3. What is the relation of the total sample mean and the mean of the sample means? How do these compare with the population mean $\mu = 0$?
4. What is the relation of the total sample variance and the sample variance of the sample means? How do these compare with the population variance $\sigma^2 = 1$?

Exercise 4.11

Statement of the Problem. Repeat Exercise 2.2, emphasizing the standard normal distribution. Answer the same set of questions you did in Exercise 2.2

Exercise 4.12

Statement of the Problem. Prepare and execute a program which will take n observations from a population which is N(10,4). From these observations, find the cumulative frequencies less than 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16. Construct a cumulative relative frequency histogram.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
n = 100	n = 500

Output. Your program should output the cumulative frequencies less than 4, 5, 6, . . . , 16.

Discussion and Questions. Since the normal distribution plays such an important role in statistical theory and applications, areas under the normal curve are readily available from tables. These tables frequently give the area under the normal curve between the mean 0 and any positive z value, where $z = (x - \mu)/\sigma$ and is referred to as a standard unit. Since the normal distribution is symmetrical about the mean μ, it is not necessary for the table to give areas for negative values of z. The purpose of this exercise is to illustrate the concept of area under the normal curve.

1. What percentage of the observations fall between 9 and 14? Compare this to the theoretical value of 66.87%.

http

2. What percentage of the observations fall between 13 and 15? Compare this to the theoretical value of 6.06%.
3. Plot the thirteen points on normal probability graph paper. Read the mean and variance off the graph. Compare all the answers to the theoretical normal distribution.

Exercise 4.13

Statement of the Problem. The normal distribution is used to approximate the binomial distribution. This choice is justifiable when the sample size n is large. Prepare and execute a program to show that this approximation is reasonable for various values of n.

Input. Test your program with the following sets of values.

Data Set 1	*Data Set 2*
p = 1/2	p = 1/2
n = 10	n = 25

Output. Your program should output the values of the binomial and normal distributions as well as the difference between the two.

Discussion and Questions. When the sample size n is large, the calculation of binomial probabilities becomes very difficult, since large powers of p and 1 - p are involved and the number of terms to be summed becomes large. It can be shown that as the sample size n increases, the binomial distribution approaches a normal distribution with mean np and variance np(1 - p). Again n represents the number of trials and p represents the probability of success on a single trial. The most readily available tables of the binomial distribution are those for n = 2 to 25. The binomial distribution has been tabulated for sample sizes from 2 to 100. If tables are not available or if the sample size is greater than 100, then we shall use the normal distribution to approximate the binomial distribution.

1. Compare the answers obtained from the normal and binomial distribution for the two data sets.
2. Does the approximation appear to get better as n increases?
3. What justification does one have in approximating a discrete distribution by a continuous distribution?

http

5

Statistical Inference—Large Sample Methods

Prerequisites

The exercises of this chapter require a basic understanding of the normal distribution, the Central Limit Theorem, point estimation, interval estimation when σ^2 is known or when the sample size is large, and tests of hypotheses when σ^2 is known or when the sample size is large.

Introduction

An interval will be called a *random interval* if at least one of its end points is a random variable. Let X be a random variable and consider the event $0 < X < 1$. This event is equivalent to $0 < X$ and $X < 1$. Now $0 < X$ can be rewritten as $1 < X + 1$ by adding one to both sides. Therefore, the interval $0 < X < 1$ can be rewritten as $X < 1 < X + 1$. The interval is a random interval, since at least one of its end points is a random variable. In this example, both of the end points are random variables. If p is the probability

$$P(X < 1 < X + 1) = P(0 < X < 1),$$

then p is precisely the probability that the random variable X satisfies $0 < X < 1$. If X is a random variable from a continuous uniform distribution, then $P(0 < X < 1) = 1$. If X is a random variable from a normal distribution with mean zero and variance 1, then $P(0 < X < 1)$ is .341.

In each of these examples, X is called a *statistic,* since it is a function of one or more random variables that does not depend upon any unknown parameter. If X were a random variable from a normal distribution with unknown mean μ and unknown variance σ^2, then $Y = (X - \mu)/\sigma$ would be a random variable but *not* a statistic, since it depends upon two unknown parameters of the normal distribution, μ and σ^2. If μ and σ^2 were known numbers, then Y would be a statistic. An important fact is that Y is N(0,1). It is quite possible that a statistic does not depend upon any unknown parameter, but the distribution of that statistic may well depend upon unknown parameters. As an example, consider the well-known statistic

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

It clearly does not depend upon any unknown parameters and it is a function of n random variables $X_1, X_2, \ldots, X_n$. If we let $X_1, X_2, \ldots, X_n$ be n outcomes of a random experiment performed n independent times under identical conditions, each having a normal distribution with known variance σ^2 and *unknown* mean μ, then $\overline{X}$ has a normal distribution with mean μ and variance σ^2/n. Clearly, the distribution of the statistic $\overline{X}$ depends upon an unknown parameter and $\sqrt{n}\,(\overline{X} - \mu)/\sigma$ is N(0,1). Then

$$P\left(-2 < \sqrt{n}\,\frac{(\overline{X} - \mu)}{\sigma} < 2\right) = .954,$$

or equivalently,

$$P\left(\overline{X} - \frac{2\sigma}{\sqrt{n}} < \mu < \overline{X} + \frac{2\sigma}{\sqrt{n}}\right) = .954.$$

Hence, we have found that the probability is .954 that the random interval $(\overline{X} - 2\sigma/\sqrt{n},\ \overline{X} + 2\sigma/\sqrt{n})$ contains the fixed (but unknown) mean μ. Each of the random variables $\overline{X} - 2\sigma/\sqrt{n}$ and $\overline{X} + 2\sigma/\sqrt{n}$ is a statistic, since σ is a known number. So far we have only discussed probability and our problem has not become statistical, since the above probability states that *prior* to the repeated independent performances of the *random* experiment, the probability is .954 that the random interval $(\overline{X} - 2\sigma/\sqrt{n},\ \overline{X} + 2\sigma/\sqrt{n})$ includes the unknown fixed parameter μ.

Suppose now we conduct the *n* random experiments and we obtain $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$; hence the sample value of $\overline{X}$ is $\overline{x}$, a known number. Clearly,

http

$\bar{x} \pm 2\sigma/\sqrt{n}$ is known since σ is known, but we cannot say that the probability is .954 that the interval $(x - 2\sigma/\sqrt{n},\ x + 2\sigma/\sqrt{n})$ includes the fixed parameter μ. Even though we cannot say the probability is .954, we rely heavily on the high degree of probability (.954) prior to the performance of the experiment to estimate the confidence we have that the interval $(\bar{x} - 2\sigma/\sqrt{n},\ x + 2\sigma/\sqrt{n})$ contains μ. The number .954 is called the *confidence coefficient*, and statistical inference of this sort is an example of *interval estimation* of a parameter. Exercise 5.1 will further illustrate this concept.

The two principle areas of statistical inference are the areas of estimation of parameters and tests of hypotheses. We have just seen an example of interval estimation of a parameter. Examples of point estimation have been seen previously for example, Exercise 4.4. The subject of tests of statistical hypotheses and estimation will be introduced through a series of exercises. For the most part, the purpose of these exercises is to acquaint the student with the basic principles of the test of a hypothesis and estimation. These are *not* an application to a practical problem, since in practice one does not test various hypotheses with the same sample. The exercises of this chapter are concerned with large and asymptotic distributions where the Central Limit Theorem will be utilized.

Exercise 5.1

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations from a population which is N(10,100). Let us pretend that the population mean is unknown. Using the 5% confidence coefficient, compute two confidence intervals for μ (the population mean). In the first, compute the confidence interval assuming σ^2 is known, and in the second, compute the confidence interval assuming σ^2 is unknown, that is, just replace σ^2 by the sample variance. Repeat the above experiment N times. Tabulate the number of times the population mean $\mu = 10$ falls within the calculated intervals.

Input. Test your program with the following sets of input values.

Data Set 1	*Data Set 2*
n = 10	n = 40
N = 100	N = 100

Output. Your program should output the number of times the population mean μ falls within the calculated intervals, the sample mean and variance for each of the N experiments, and the total sample mean and variance.

Discussion and Questions. The purpose of this exercise is threefold. First, the exercise is to show that in repeated sampling, the population mean will fall in 95% = (100 - 5)% of the calculated 5% confidence intervals. Second, when the population variance is unknown and when the sample size is large, we can rely heavily on the high degree of probability (confidence coefficient) established prior to the running of the experiment. Third, as the sample size increases within each

experiment, the better are the point estimates $\bar{x}$ and s^2 of μ and σ^2, respectively. As a guideline, n should be greater than 30 for this approximation to be valid.

1. For each data set and σ^2 known, how does the tabulated number of times the population mean μ falls within the calculated interval compare with the theoretical result, that is, N × .95?
2. For each data set and σ^2 unknown, how does the tabulated number of times the population mean μ falls within the calculated interval compare with the theoretical result, that is, N × .95?
3. How do the answers to Questions 1 and 2 compare? What is the consequence of increasing the sample size without changing the confidence coefficient?
4. What would be the effect if we reduced the confidence coefficient to 1%?
5. What is the consequence of using zero percent confidence coefficient?
6. What is the advantage of having a large sample in determining the confidence interval of a population mean? Illustrate your answer with the output of your computer experiment.
7. Why do the lengths of the confidence intervals of a population mean change from sample to sample even if the sample size remains the same?
8. Regardless of the sample size, one can obtain a confidence interval. Why does one prefer a large sample? Comment about your point estimates of μ and σ^2.

Exercise 5.2

40

Statement of the Problem. Prepare and execute a program that will take two random samples of size n_1 and n_2 from populations which are N(5,1) and N(10,1), respectively. Let us pretend that the population means are unknown. Using the 5% confidence coefficient, compute two confidence intervals for $\mu_2 - \mu_1$ (the difference between the two population means). In the first, compute the confidence interval assuming σ_1^2 and σ_2^2 are known; and in the second, compute the confidence interval assuming σ_1^2 and σ_2^2 are unknown. Repeat the above experiment N times. Tabulate the number of times the actual difference of the two population means, $\mu_2 - \mu_1 = 5$, falls within the calculated intervals.

Input. Test your program with the following sets of input values.

Data Set 1	*Data Set 2*
$n_1 = 20$, $n_2 = 10$	$n_1 = 40$, $n_2 = 30$
$N = 100$	$N = 100$

Output. Your program should output the number of times the difference of the two population means $\mu_2 - \mu_1$ falls within the calculated intervals, the sample mean and variance for each of the N experiments, and the total sample mean and variance.

Discussion and Questions. The purpose of this exercise is threefold. First, the exercise is to show that in repeated sampling, the difference of the two population means will fall in 95% = (100 - 5)% of the calculated 5% confidence intervals.

http

Second, when the population variances are unknown, we can rely heavily on the high degree of probability (confidence coefficient) established prior to running the experiment only when the sample sizes are large. As a guideline, n_1 and n_2 each should be greater than 30 for the approximation to be valid. Third, as the sample size increases within each experiment, the better are the point estimates $\bar{x}$ and s^2 of μ and σ^2, respectively.

1. For each data set and σ_1^2, σ_2^2 known, how does the tabulated number of times the difference of the two population means $\mu_2 - \mu_1$ falls within the calculated interval compare with the theoretical result, that is, N X .95?
2. For each data set and σ_1^2, σ_2^2 assumed unknown, how does the tabulated number of times the difference of the two population means $\mu_2 - \mu_1$ falls within the calculated interval compare with the theoretical result, that is, N X .95?
3. How do the answers to Questions 1 and 2 compare? What is the consequence of increasing the sample size without changing the confidence coefficient?
4. What would be the effect if we reduced the confidence coefficient to 1%?
5. What is the consequence of using zero percent confidence coefficient?
6. What is the advantage of having a large sample in determining the confidence interval for the difference of two population means? Illustrate your answer with the output of your computer experiment.
7. Why do the lengths of the confidence intervals of the difference of two population means change from sample to sample even if the sample size remains the same?
8. Regardless of the sample size, one can obtain a confidence interval. Why does one prefer a large sample? Comment about your point estimates of μ and σ^2.

Exercise 5.3

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations from a population which is N(50,100). Let us pretend that the population mean is unknown. Using the 5% significance level, test the hypothesis that the population mean μ is equal to μ_0 (a constant). This exercise will be a two-tailed test.

Input. Test your program with the following set of input values.

Data Set 1

μ_0 = 40, 49, 50, 51, and 60
n = 16

Output. Your program should output the sample mean, the sample variance, and the standardized z-score for each value μ_0.

Discussion and Questions. In this exercise, it is possible to determine whether a conclusion concerning $\mu = \mu_0$ is right or wrong, since the population mean is actually known to be 50. Therefore, for each of the five cases of μ_0 stated in Data Set 1, it is

http

possible to determine whether the conclusion is correct, a Type-I error made, or a Type-II error has been made. This exercise is intended to show that a Type-II error is likely to be committed if the hypothetical population mean μ_0 is close to the true population mean 50.

1. For the hypothesis $\mu = \mu_0 = 50$, write a complete report including the following four important elements of a statistical test: null hypothesis, test statistic, rejection region and alternative hypothesis. For the rest of the hypotheses, just compute and state the conclusions; that is, is the conclusion correct or has a Type-I or a Type-II error been made?
2. From your output, is a Type-II error more likely to be committed when the hypothetical population mean μ_0 is close to the true population mean 50?
3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II error?
4. If a hypothesis is already rejected by a sample of 16 observations, is it likely to be rejected or accepted by a sample of 100 observations? Why?
5. If one replaced the known value of σ^2 by the sample variance s^2, would this change the conclusion concerning the five null hypotheses stated in Data Set 1? Is this procedure valid? Why?
6. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
7. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?

Exercise 5.4

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations from a population which is N(50,100). Let us pretend that the population mean is unknown. Using the 1% significance level, test the hypothesis that the population mean μ is equal to μ_0 (a constant). This will be a two-tailed test.

Input. Test your program with the following set of input values.

Data Set 1

μ_0 = 40, 49, 50, 51, and 60
n = 500

Output. Your program should output the sample mean, the sample variance and the standardized z-score for each value μ_0.

Discussion and Questions. In this exercise, as well as in Exercise 5.3, it is possible to determine whether a conclusion concerning $\mu = \mu_0$ is right or wrong, since the population mean is actually known to be 50. Therefore, it is possible to determine whether the conclusion is correct, a Type-I error is made, or a Type-II error is made. This exercise is intended to show that the probability of committing both Type-I and Type-II errors can be reduced by increasing the sample size.

http

1. For the hypothesis $\mu = \mu_0 = 50$, write a complete report including the following four important elements of a statistical test: null hypothesis, test statistic, rejection region and alternative hypothesis. For the rest of the hypotheses, just compute and state the conclusions; that is, is the conclusion correct, or has a Type-I or a Type-II error been made?
2. Does one need a large or small sample to reject a false hypothesis which is very close to the true one? Why? (Relate your answer to the output of your program.)
3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II error?
4. When one tests a hypothesis with the same significance level, is a large sample or a small sample more likely to cause the rejection of the hypothesis? Why?
5. If one replaced the known value of σ^2 by the sample variance s^2, would this change the conclusion concerning the five null hypotheses stated in Data Set 1? Is this procedure valid? Why?
6. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
7. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?

Exercise 5.5

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations from a population which is N(50,100). Let us pretend that the population mean is unknown. Using both the 1% and 5% significance levels, test the hypothesis that the population mean μ is equal to μ_0 (a constant). This exercise will be a two-tailed test.

Input. Test your program with the following set of input values.

Data Set 1

$\mu_0 = 49, 50$
$n = 25$

Output. Your program should output the sample mean, the sample variance, and the standardized z-score for each value μ_0.

Discussion and Questions. In this exercise, as well as in Exercises 5.3 and 5.4, it is possible to determine whether a conclusion concerning $\mu = \mu_0$ is right or wrong, since the population mean is actually known to be 50. Therefore, it is possible to determine whether the conclusion is correct, a Type-I error has been made, or a Type-II error has been made. This exercise is intended to show that a change in significance level without a change in sample size gains something and also loses something. The use of the 5% significance level is more likely to lead to a Type-I error and less likely to a Type-II error than the use of the 1% significance level for the same sample size.

http

1. For the hypothesis $\mu = \mu_0 = 50$, write a complete report including the following four important elements of a statistical test: null hypothesis, test statistic, rejection region and alternative hypothesis. For the rest of the hypotheses, just compute and state the conclusions; that is, is the conclusion correct, or has a Type-I or a Type-II error been made?
2. From your output, what is the effect of changing the significance level without a change in the sample size?
3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II error?
4. If a hypothesis is rejected using a 5% significance level, is it likely to be rejected using a 1% significance level? Why? Is this a valid question? Why?
5. If one replaced the known value of σ^2 by the sample variance s^2, would this change the conclusion concerning the five null hypotheses stated in Data Set 1? Is this procedure valid? Why?
6. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
7. In practice, is it valid to test a hypothesis using both the 1% and 5% significance level? Why?
8. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?

Exercise 5.6

Statement of the Problem. Prepare and execute a program that will take two random samples of size n_1 and n_2 from populations which are N(50,100) and N(50,25), respectively. Let us pretend that the population means μ_1, μ_2 are unknown. Using the 5% significance level, test the hypothesis that the difference between the two population means $\mu_2 - \mu_1$ is equal to d (a constant). This exercise is a two-tailed test.

Input. Test your program with the following set of input values.

Data Set 1

$d = -10, -1, 0, 1,$ and 10
$n_1 = 20, n_2 = 25$

Output. Your program should output the sample means, the sample variances, and the standardized z-score for each value of d.

Discussion and Questions. In this exercise, it is possible to determine whether a conclusion concerning $\mu_2 - \mu_1 = d$ is right or wrong, since the difference between the two population means is actually known to be 0. Therefore, for each of the five cases of d stated in Data Set 1, it is possible to determine whether the conclusion is correct, a Type-I error has been made, or a Type-II error has been made. This exercise is intended to show that a Type-II error is likely to be committed if the hypothetical difference of the two population means *d* is close to the true difference between the two population means 0.

http

1. For the hypothesis $\mu_2 - \mu_1 = d$, write a complete report including the following four important elements of a statistical test: null hypothesis, test statistic, rejection region and alternative hypothesis. For the rest of the hypotheses, just compute and state the conclusions; that is, is the conclusion correct or has a Type-I or a Type-II error been made?
2. From your output, is a Type-II error more likely to be committed when the hypothetical population mean μ_0 is close to the true population mean 50?
3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II error?
4. If a hypothesis is already rejected by a sample of $n_1 = 20$ and $n_2 = 25$ observations, is it likely to be rejected or accepted by a sample of $n_1 = n_2 = 100$ observations? Why?
5. If one replaced the known value of σ^2 by the sample variance s^2, would this change the conclusion concerning the five null hypotheses stated in Data Set 1? Is this procedure valid? Why?
6. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
7. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?

Exercise 5.7

Statement of the Problem. Prepare and execute a program that will take two random samples of size n_1 and n_2 from populations which are N(50,100) and N(50,25), respectively. Let us pretend that the population means μ_1, μ_2 are unknown. Using the 1% significance level, test the hypothesis that the difference between the two population means $\mu_2 - \mu_1$ is equal to d (a constant). This exercise is a two-tailed test.

Input. Test your program with the following set of input values.

Data Set 1

$d = -10, -1, 0, 1,$ and 10
$n_1 = 200, \quad n_2 = 500$

Output. Your program should output the sample means, the sample variances, and the standardized z-scores for each value of d.

Discussion and Questions. In this exercise, as well as in Exercise 5.6, it is possible to determine whether a conclusion concerning $\mu_2 - \mu_1 = d$ is right or wrong, since the difference between the two population means is actually known to be 0. Therefore, it is possible to determine whether the conclusion is correct, a Type-I error has been made, or a Type-II error has been made. This exercise is intended to show that the probability of committing both Type-I and Type-II errors can be reduced at the same time by increasing the sample size.

http

1. For the hypothesis $\mu_2 - \mu_1 = d$, write a complete report including the following four important elements of a statistical test: null hypothesis, test statistic, rejection region and alternative hypothesis. For the rest of the hypotheses, just compute and state the conclusions; that is, is the conclusion correct or has a Type-I or a Type-II error been made?
2. Does one need a large or small sample to reject a false hypothesis which is very close to the true one? Why? (Relate your answer to the output of your program.)
3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II error?
4. When one tests a hypothesis with the same significance level, is a large sample or a small sample more likely to cause the rejection of the hypothesis? Why?
5. If one replaced the known value of σ^2 by the sample variance s^2, would this change the conclusion concerning the five null hypotheses stated in Data Set 1? Is this procedure valid? Why?
6. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
7. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?

Exercise 5.8

Statement of the Problem. Prepare and execute a program that will take two random samples, n_1 and n_2, from populations which are N(50,100) and N(50,25), respectively. Let us pretend that the population means μ_1, μ_2 are unknown. Using both the 1% and the 5% significance levels, test the hypothesis that the difference between the two population means $\mu_2 - \mu_1$ is equal to d (a constant). This exercise is a two-tailed test.

Input. Test your program with the following set of input values.

Data Set 1

$d = -1, 0, 1$
$n_1 = 100, \quad n_2 = 25$

Output. Your program should output the sample means, the sample variances, and the standardized z-scores for each value of d.

Discussion and Questions. In this exercise, as well as in Exercises 5.6 and 5.7, it is possible to determine whether a conclusion concerning $\mu_2 - \mu_1 = d$ is right or wrong, since the difference between the two population means is actually known to be 0. Therefore, it is possible to determine whether the conclusion is correct, a Type-I error has been made, or a Type-II error has been made. This exercise is intended to show that a change in sample size gains something and also loses something. The use of the 5% significance level is more likely to lead to a Type-I error and less likely to a Type-II error than the use of the 1% significance level.

http

1. For the hypothesis $\mu_2 - \mu_1 = d$, write a complete report including the following four important elements of a statistical test: null hypothesis, test statistic, rejection region and alternative hypothesis. For the rest of the hypotheses, just compute and state the conclusions; that is, is the conclusion correct or has a Type-I or a Type-II error been made?
2. From your output, what is the effect of changing the significance level without a change in the sample size?
3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II error?
4. If a hypothesis is rejected by using a 5% significance level, is it likely to be rejected using a 1% significance level? Why? Is this a valid question? Why?
5. If one replaced the known value of σ^2 by the sample variance s^2, would this change the conclusion concerning the five null hypotheses stated in Data Set 1? Is this procedure valid? Why?
6. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
7. In practice, is it valid to test a hypothesis using both the 1% and 5% significance level? Why?
8. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?

http

6

Statistical Inference—Exact Distributions

Prerequisites

The exercises of this chapter require a basic understanding of the chi-square distribution, the t/distribution, the F/distribution, point estimation, interval estimation, and tests of hypotheses.

Introduction

The examples and exercises contained in Chapter 5 were concerned with large sample methods for making inferences concerning population means and the difference between two population means. In this chapter we shall consider inferential procedures which depend upon exact distributions and are closely related to the large sample methods.

In the introduction of Chapter 5, we noted that if we let $X_1, X_2, \ldots, X_n$ be n outcomes of a random experiment performed n independent times under identical conditions, each having a normal distribution with mean μ and variance σ^2, then

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

has a normal distribution with mean μ and variance σ^2/n, and $z = \sqrt{n}\ (\overline{X} - \mu)/\sigma$ has a normal distribution with mean zero and variance one. From Exercise 4.10, we saw that the distribution of z was approximately a normal distribution with mean zero and variance one in repeated sampling when n is large (Central Limit Theorem). It was also noted that the original population did not need to be normal for the Central Limit Theorem; that is, each of the random variables $X_1, X_2, \ldots, X_n$ need *not* have a normal distribution. From Exercise 4.4, we saw that the point estimate

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{(n - 1)}$$

approaches σ^2 as the sample size increases. It should appear quite reasonable from the exercises in Chapter 5 that the Central Limit Theorem holds if σ is known, or that a good estimate, s, is available and is based upon a reasonable large sample. It is recommended as a guideline that n be greater than or equal to 30. Therefore, we have taken care of the case where $n \geqslant 30$, but the problem remains as to what do we do when n is small, say n = 5.

Before proceeding to the exercises, we will review briefly the chi-square distribution, the t distribution, and the F distribution. Each of these plays an important role in statistical inference involving exact distribution.

It can be shown mathematically that if z follows a normal distribution with mean zero and variance one, then z^2 has a *chi-square* distribution with one degree of freedom. Also, it may be shown mathematically that the quantity

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{\sigma^2}$$

follows a chi-square distribution with n - 1 degrees of freedom.

In order to discuss the t distribution, we must assume that $\overline{X}$ and s^2 are independent. If $\overline{X}$ and s^2 are independent (which they are), then $z^2 = n(\overline{X} - \mu)^2/\sigma^2$ and $(n - 1)^2/\sigma^2$ are independent. In accordance with the previous fact, define the random variable T as follows:

$$T = \frac{[\sqrt{n}(\overline{X} - \mu)]/\sigma}{\sqrt{ns^2/[\sigma^2(n - 1)]}} = \frac{\sqrt{n}(\overline{X} - \mu)}{s}.$$

We say that T has a t distribution with n - 1 degrees of freedom. It should be obvious from the above construction of T, that the numerator has a normal distribution with mean zero and variance one, while the denominator is the square root of an independent chi-square distribution divided by n - 1, its degrees of freedom. Both the chi-square and t distributions will be used to establish confidence intervals and

http

test of hypotheses for the population variance and population means, when the sample size is small.

The last distribution to be discussed here is the *F* distribution. Let X and Y denote independent random variables that are $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Consider a random sample $X_1, X_2, \ldots, X_{n_1}$, of size $n_1 \geqslant 2$ from the distribution of X and a random sample $Y_1, Y_2, \ldots, Y_{n_2}$, of size $n_2 \geqslant 2$ from the independent distribution of *Y*. Note that n_1 does not need to be equal to n_2. Let X_1 and $\overline{X}_2$ denote the sample means and s_1^2 and s_2^2 the sample variances. The two independent random variables $(n_1 - 1)s_1^2/\sigma_1^2$ and $(n_2 - 1)s_2^2/\sigma_2^2$ have chi-square distributions with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, respectively. The random variable F will be defined as follows:

$$F = \frac{ns_1^2/[\sigma_1^2(n_1 - 1)]}{ns_2^2/[\sigma_2^2(n_2 - 1)]}$$

We say that F has an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. The F distribution will be used to establish confidence intervals and tests of hypotheses for two population variances.

Exercise 6.1

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations from a population which is N(50,100). Let us pretend that the population mean and variance are unknown. Using the 5% confidence coefficient, compute a confidence interval for μ (the population mean). Repeat the above experiment N times. Tabulate the number of times the population mean $\mu = 50$ falls within the calculated intervals.

Input. Test your program with the following sets of input values.

Data Set 1	*Data Set 2*
n = 10	n = 40
N = 100	N = 100

Output. Your program should output the number of times the population mean μ falls within the calculated intervals, the sample mean variance for each of the N experiments, and the total sample mean and variance.

Discussion and Questions. In this exercise we are concerned with the problem of finding a confidence interval for the population mean μ of a normal distribution when we are *not* so fortunate as to know the population variance σ^2. Since we are assuming that the population variance σ^2 is unknown, we are going to utilize that statistic χ^2 defined in the introduction. As in Exercise 5.1, the purpose of this exercise is to show that in repeated sampling, the population mean will fall in

http

95% = (100 - 5)% of the calculated 5% confidence intervals. Even though the population variance is unknown, we can rely on the high degree of probability (confidence coefficient) established prior to running the experiment, since we have an exact distribution, namely, the t distribution.

1. For each data set and σ^2 unknown, how does the tabulated number of times the population mean μ falls within the calculated interval compare with the theoretical number, that is, N X .95?
2. What is the consequence of increasing the sample size without changing the confidence coefficient?
3. What would be the effect if we reduced the confidence coefficient to 1%?
4. What is the consequence of using zero-percent confidence coefficient?
5. What is the advantage of having a large sample in determining the confidence interval of a population mean? Illustrate your answer with the output of your computer experiment.
6. Why do the lengths of the confidence intervals of a population mean change from sample to sample even when the sample size remains the same?
7. Regardless of the sample size, one can obtain a confidence interval. Why does one prefer a large sample? Comment about your point estimates of μ and σ^2.
8. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?

Exercise 6.2

Statement of the Problem. Prepare and execute a program that will take two random samples of size n_1 and n_2 from populations which are N(5,1) and N(10,1), respectively. Let us pretend that the population means and variances are unknown. Using the 5% confidence coefficient, compute a confidence interval for $\mu_2 - \mu_1$ (the difference of the two population means). Repeat the above experiment N times. Tabulate the number of times the difference of the two population means $\mu_2 - \mu_1 = 5$ falls within the calculated intervals.

Input. Test your program with the following sets of input values.

Data Set 1	*Data Set 2*
$n_1 = 20$, $n_2 = 10$	$n_1 = 40$, $n_2 = 30$
$N = 100$	$N = 100$

Output. Your program should output the number of times the difference between the two population means $\mu_2 - \mu_1$ falls within the calculated intervals, the sample mean and variance for each of the N experiments, and the total sample mean and variance.

Discussion and Questions. In this exercise we are concerned with the problem of finding a confidence interval for the difference $\mu_2 - \mu_1$ between the population means of two independent normal distributions when the distributions have the

http

same, but unknown, population variance σ^2. In the Introduction of Chapter 5, we constructed the random variable T when we were concerned with only one population mean. How is the random variable T constructed for two population means? The random variable T for two population means may be constructed as follows: let $x_1, x_2, \ldots, x_{n_1}$ and $y_1, y_2, \ldots, y_{n_2}$ denote independent random samples from two independent distributions having respectively the probability density functions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$. Let $\bar{X}$ and $\bar{Y}$ denote the sample means and s_1^2 and s_2^2 the sample variances. Thus $\bar{X}$ and $\bar{Y}$ are normally and independently distributed with means μ_1 and μ_2 and variances σ^2/n_1 and σ^2/n_2, respectively. Accordingly, their difference $\bar{X} - \bar{Y}$ is normally distributed with mean $\mu_1 - \mu_2$ and variance $\sigma^2/n_1 + \sigma^2/n_2$. Thus the random variable

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2/n_1 + \sigma^2/n_2}}$$

is normally distributed with mean zero and variance one. This random variable will serve as the numerator of a T random variable.

$$\frac{(n_1 - 1)s_1^2}{\sigma^2} \quad \text{and} \quad \frac{(n_2 - 1)s_2^2}{\sigma^2}$$

have independent chi-square distributions with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, respectively, so that

$$\frac{[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]}{\sigma^2}$$

has a chi-square distribution with $(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$ degrees of freedom ($n_1 + n_2 - 2 > 0$). In accordance with the above statements, define the random variable T as follows:

$$T = \frac{\dfrac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(1/n_1 + 1/n_2)}}}{\sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{\sigma^2(n_1 + n_2 - 2)}}} = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

We say that T has a t distribution with $n_1 + n_2 - 2$ degrees of freedom. It should be obvious from the above construction of T that the numerator has a normal distribution with mean zero and variance one, and the denominator is the square root of an independent chi-square distribution divided by $n_1 + n_2 - 2$, its degrees of freedom as before. As in Exercise 6.2, the purpose of this exercise is to show that in repeated sampling, the difference of the two population means will fall in 95% = (100 - 5)% of the calculated 5% confidence intervals. Even though the population variances are unknown, we can rely on the high degree of probability (confidence coefficient) established prior to running the experiment, since we have an exact distribution, namely, the t distribution.

1. For each data set and σ^2 unknown, how does the tabulated number of times the difference of the two population means $\mu_2 - \mu_1$ falls within the calculated interval compare with the theoretical result, that is, N × .95?
2. What is the consequence of increasing the sample size without changing the confidence coefficient?

3. What would be the effect if we reduced the confidence coefficient to 1%?
4. What is the consequence of using zero percent confidence coefficient?
5. What is the advantage of having a large sample in determining the confidence interval for the difference of two population means? Illustrate this with the output of your computer experiment.
6. Why do the lengths of the confidence intervals of the difference of two population means change from sample to sample even when the sample size remains the same?
7. Regardless of the sample size, one can obtain a confidence interval. Why does one prefer a large sample? Comment about your point estimates of μ and σ^2.
8. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?

Exercise 6.3

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations from a population which is N(50,100). Let us pretend that the population mean and variance are unknown. Using the 5% significance level, test the hypothesis that the population mean μ is equal to μ_0 (a constant). This exercise will be a two-tailed test.

Input. Test your program with the following set of input values.

Data Set 1

μ_0 = 40, 49, 50, 51, and 60
n = 10

Output. Your program should output the sample mean, the sample variance, and the value of T for each value μ_0.

Discussion and Questions. In this exercise, it is possible to determine whether a conclusion concerning $\mu = \mu_0$ is right or wrong, since the population mean is actually known to be 50. Therefore, for each of the five cases of μ_0 stated in Data Set 1, it is possible to determine whether the conclusion is correct, a Type-I error has been made, or a Type-II error has been made. This exercise is intended to show that a Type-II error is likely to be committed if the hypothetical population mean μ_0 is close to the true population mean 50. The only difference between this exercise and Exercise 5.3 is that the t distribution will be utilized, since the population variance is unknown and the sample size is small.

1. For the hypothesis $\mu = \mu_0 = 50$, write a complete report including the following four important elements of a statistical test: null hypothesis, test statistic, rejection region and alternative hypothesis. For the rest of the hypotheses, just compute and state the conclusions; that is, is the decision correct or has a Type I or a Type II error been made?
2. From your output, is a Type II error more likely to be committed when the hypothetical population mean μ_0 is close to the true population mean 50?

http

3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II errors?
4. If a hypothesis is already rejected by a sample of 10 observations, is it likely to be rejected or accepted by a sample of 100 observations? Why?
5. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
6. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?
7. What would be the effect of changing the significance level from 5% to 1% along with increasing the sample size from 10 to 25?

Exercise 6.4

Statement of the Problem. Prepare and execute a program that will take two random samples of size n_1 and n_2 from populations which are $N(\mu_1, 100)$ and $N(\mu_2, 100)$ respectively. Let us pretend that the population means μ_1, μ_2 are unknown and the population variance σ^2 is unknown. Using the 5% significance level, test the hypothesis that the difference between the two population means $\mu_2 - \mu_1$ is equal to d (a constant). This exercise is a two-tailed test.

Input. Test your program with the following sets of input values.

Data Set 1	*Data Set 2*	*Data Set 3*
$\mu_1 = \mu_2 = 50$	$\mu_1 = 49,\ \mu_2 = 51$	$\mu_1 = 30,\ \mu_2 = 70$
$n_1 = 6,\ n_2 = 7$	$n_1 = 6,\ n_2 = 7$	$n_1 = 6,\ n_2 = 7$
$d = 0$	$d = 0$	$d = 0$

Output. Your program should output the sample means, the sample variances, and the value of T for each data set.

Discussion and Questions. In this exercise, we are going to use the random variable T to test the hypothesis $\mu_2 - \mu_1 = d = 0$ (or $\mu_1 = \mu_2$). When dealing with Data Set 2, it may be easier to subtract 1 from each of the six observations of Sample 1 (Data Set 1) and add 1 to each of the seven observations of Sample 2 (Data Set 1). Similarly in Data Set 3, it may be easier to subtract 20 from each of the six observations of Sample 1 (Data Set 1) and add 20 to each of the seven observations of Sample 2 (Data Set 2). Using these three data sets, it is possible to determine whether a conclusion concerning $\mu_2 - \mu_1 = 0$ (or $\mu_1 = \mu_2$) is right or wrong. Therefore, for each of the three data sets, it is possible to determine whether the conclusion is correct, a Type-I error has been made, or a Type-II error has been made.

The purpose of this exercise is threefold. First, a Type-II error is likely to be committed if the hypothetical population mean is close to the true population mean. Secondly, a Type-II error is likely to be made if the sample sizes are small and the hypothesis not too far wrong. Finally, a false hypothesis can be rejected with relatively small samples if the hypothesis is far enough from the truth.

http

1. From Data Set 1, state whether your conclusion is correct or a Type-I error has been committed. Is it possible for a Type-II error to be committed? Why?
2. From Data Sets 2 and 3, state whether your conclusion is correct or a Type-II error has been committed. Is it possible for a Type-I error to be committed? Why?
3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II errors?
4. If a hypothesis is already rejected by a sample of $n_1 = 6$, $n_2 = 7$ observations, is it likely to be rejected by a sample of $n_1 = n_2 = 100$ observations? Why?
5. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
6. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?
7. What would be the effect of changing the significance level from 5% to 1% and increasing the sample size from $n_1 = 6$, $n_2 = 7$ to $n_1 = n_2 = 50$?
8. What happens if σ_1^2 does not equal σ_2^2? What effect would this have on the statistic T?

Exercise 6.5

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations from a population which is N(10,100). Let us pretend that the population mean and variance are unknown. Using the 5% confidence coefficient, compute a confidence interval for σ^2 (the population variance). Repeat the above experiment N times. Tabulate the number of times the population variance $\sigma^2 = 100$ falls within the calculated intervals.

Input. Test your program with the following sets of input values.

Data Set 1	*Data Set 2*
n = 10	n = 50
N = 100	N = 100

Output. Your program should output the number of times the population variance falls within the calculated intervals, the sample variance and the χ^2 statistic for each of the N experiments, the total sample variance, and the mean and variance for the N computed values of χ^2.

Discussion and Questions. In this exercise we are concerned with the problem of finding a confidence interval for the population variance σ^2. We are going to utilize the χ^2 statistic defined in the introduction. The purpose of this exercise is twofold. First, we intend to show that in repeated sampling, the population variance will fall in 95% = (100 − 5)% of the calculated 5% confidence intervals. Secondly since the mean of the random variable χ^2 is n − 1 and the variance of the random variable χ^2 is 2(n − 1), we will show that the sampling distribution of

http

$$\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

has a mean approximately equal to $n - 1$ and variance approximately equal to $2(n - 1)$.

1. For each data set, how does the tabulated number of times the population variance σ^2 falls within the calculated interval compare with the theoretical result, that is, N X .95?
2. What is the consequence of increasing the sample size without changing the confidence coefficient?
3. What would be the effect if we reduced the confidence coefficient to 1%?
4. What is the consequence of using zero-percent confidence coefficient?
5. What is the advantage of having a large sample in determining a confidence interval for the population variance? Illustrate your answer with the output of your computer experiment.
6. Why do the lengths of the confidence intervals of a population variance change from sample to sample even when the sample size remains the same?
7. Regardless of the sample size, one can obtain a confidence interval. Why does one prefer a large sample? Comment about your point estimate of σ^2.
8. What distribution does the statistic $(n - 1)s^2/\sigma^2$ follow? ($\sigma^2 = 100$). How do your computed mean and variance of this distribution compare with the theoretical mean and variance? Compare the two data sets.
9. What distribution would the statistic $(n - 1)s^2/\sigma^2$ follow if n were equal to 2000? ($\sigma^2 = 100$.)
10. What distribution would the statistic $(n - 1)s^2/\sigma^2$ follow if 10 were added to each of the observations in the population before the sampling experiment was carried out?

Exercise 6.6

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations from a population which is N(50,100). Pretend that the population mean and variance are unknown. Using the 5% significance level, test the hypothesis that the population variance σ^2 is equal to σ_0^2 (a constant). This exercise is a two-tailed test.

Input. Test your program with the following set of input values.

Data Set 1

σ_0^2 = 10, 99, 100, 101, and 1000
n = 8

http

Output. Your program should output the sample variance and the value of χ^2 for each value σ_0^2.

Discussion and Questions. In this exercise it is possible to determine whether a conclusion concerning $\sigma^2 = \sigma_0^2$ is right or wrong, since the population variance is actually known to be 100. Therefore, for each of the five cases of σ_0^2 stated in Data Set 1, it is possible to determine whether the conclusion is correct, a Type-1 error has been made, or a Type-II error has been made. This exercise is intended to show that a Type-II error is likely to be committed if the hypothetical population variance σ_0^2 is close to the true population variance $\sigma^2 = 100$.

1. For the hypothesis $\sigma^2 = \sigma_0^2 = 100$, write a complete report including the following four important elements of a statistical test: null hypothesis, test statistic, rejection region, and alternative hypothesis. For the rest of the hypotheses, just compute and state the conclusions. That is, is the conclusion correct or has a Type-I or a Type-II error been made?
2. From your output, is a Type-II error more likely to be committed when the hypothetical population variance σ_0^2 is close to the true population variance 100?
3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II errors?
4. If a hypothesis is already rejected by a sample of 8 observations, is it likely to be rejected or accepted by a sample of 100 observations? Why?
5. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
6. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?
7. What would be the effect of changing the significance level from 5% to 1% and increasing the sample size from 8 to 25?
8. What is the effect of adding 10 to each observation of the random sample?
9. What is the effect of multiplying each observation of the random sample by a constant?

Exercise 6.7

Statement of the Problem. Prepare and execute a program that will take two random samples of size n_1 and n_2 from populations which are $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Let us pretend that both the population means and population variances are unknown. Using the 5% significance level, test the hypothesis that the difference between the two population variances $\sigma_2^2 - \sigma_1^2$ is equal to c (a constant). This exercise is a two-tailed test.

Input. Test your program with the following sets of input values.

http

Data Set 1	*Data Set 2*	*Data Set 3*
$\mu_1 = 40,\ \mu_2 = 60$	$\mu_1 = 40,\ \mu_2 = 60$	$\mu_1 = 40,\ \mu_2 = 60$
$\sigma_1^2 = \sigma_2{}^2 = 100$	$\sigma_1^2 = 100, \sigma_2{}^2 = 121$	$\sigma_1^2 = 100, \sigma_2{}^2 = 2500$
$n_1 = 8, n_2 = 9$	$n_1 = 8, n_2 = 9$	$n_1 = 8, n_2 = 9$
$c = 0$	$c = 0$	$c = 0$

Output. Your program should output the sample variances and the F statistic for each data set.

Discussion and Questions. In this exercise, we are going to use the random variable F described in the introduction to test the hypothesis

$$\sigma_1^2 - \sigma_2^2 = c = 0 \text{ (or } \sigma_1^2 = \sigma_2^2).$$

When dealing with Data Set 2, it is easier to multiply each observation of the second sample (Data Set 1) by 1.1 and recalculate $s_2{}^2$. Similarly, in Data Set 3, it is easier to multiply each observation of the second sample (Data Set 1) by 5 and recalculate s_2^2. Using these three data sets, it is possible to determine whether a conclusion concerning

$$\sigma_1^2 - \sigma_2^2 = 0 \quad (\text{or } \sigma_1^2 = \sigma_2^2)$$

is right or wrong. Therefore, for each of the three data sets, it is possible to determine whether the decision is correct, a Type-I error has been made, or a Type-II error has been made. The purpose of this exercise is threefold. First, a Type-II error is likely to be committed if the hypothetical population mean is close to the true population mean. Secondly, a Type-II error is likely to be made if the sample sizes are small and the hypothesis is not too far wrong. Finally, a false hypothesis can be rejected with relatively small samples if the hypothesis is far enough from the truth.

1. From Data Set 1, state whether your conclusion is correct or a Type-I error has been committed. Is it possible for a Type-II error to be committed? Why?
2. From Data Sets 2 and 3, state whether your conclusion is correct or a Type-II error has been committed. Is it possible for a Type-I error to be committed? Why?
3. What would be the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II errors?
4. If a hypothesis is already rejected by a sample of $n_1 = 6, n_2 = 7$ observations, is it likely to be rejected by a sample of $n_1 = n_2 = 100$ observations? Why?
5. Regardless of the sample size, one can test a hypothesis. Why does one prefer a large sample?
6. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?

http

7. What would be the effect of changing the significance level from 5% to 1% and increasing the sample size from $n_1 = 8$, $n_2 = 9$ to $n_1 = n_2 = 50$?
8. What happens if σ_1^2 does not equal σ_2^2? What effect would this have on the statistic F?
9. If the hypothesis is accepted using Data Set 1, find the pooled estimate of the variance σ^2 which is common to both populations. What is the number of degrees of freedom of this pooled variance σ_p^2?
10. Combine the observations of the two samples of Data Set 1 together to form a single sample of 17 observations. Find the sample variance s^2 of this sample. Compare this s^2 with s_p^2 obtained in Question 9. Which quantity is greater? Why? Under what condition is s_p^2 the approximate estimate of σ^2? Why? Under what condition is s^2 the more desirable estimate of σ^2? Why?

http

7

Linear Regression and Correlation

Prerequisites

The exercises of this chapter require a basic understanding of simple linear regression and statistical inference in simple linear regression.

Introduction

The simplest relationship between two variables is a linear one, namely, $Y = \alpha + \beta X$, where α and β are unknown parameters. In this linear relationship, α denotes the intercept of the line and β denotes the slope. The practical situation is that we have n pairs of sample observations on X, Y, which can be represented on a scatter diagram. From this scatter diagram, we may be inclined to draw a straight line to model the observed relationship indicated by the scatter diagram. Clearly, if we draw a line $\alpha + \beta X$ through a scatter diagram, it will not provide an *exact* explanation of the true relationship between Y and X. This lack of fit does not condemn the model. In practice, we judge the "goodness" of a model by whether it predicts with an error that is negligible. Because of a need for accuracy or a small error of

http

prediction, we resort to a probabilistic model for Y. The following is such a model:

$$Y_i = \alpha + \beta X_i + u_i,\ i = 1, 2, \ldots, n$$
$$E(u_i) = 0 \text{ for all } i$$
$$E(u_i u_j) = \begin{cases} 0 & \text{for } i \neq j;\ i, j = 1, 2, \ldots, n \\ \sigma_u^2 & \text{for } i = j;\ i, j = 1, 2, \ldots, n \end{cases}$$

where α, β, and σ_u^2 are unknown parameters and E denotes the expected value or mean of a random variable. Now u is a random variable representing the variability about the linear relationship, $\alpha + \beta X$. We wish to estimate α, β, and σ_u^2 statistically on the basis of our sample observations on X and Y. We may also wish to test hypotheses or establish confidence intervals about these parameters. In order to accomplish this, we estimate the true line $Y = \alpha + \beta X$ by $\hat{Y} = \hat{\alpha} + \hat{\beta} X$, where

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i (Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

and

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

The following exercises are concerned with the above model, or set of assumptions, and the problem of making statistical inferences concerning the model.

Exercise 7.1

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations (u_i) from a population which is N(0,1). From these observations, find $\bar{y}$, $\hat{\alpha}$, $\hat{\beta}$, $\bar{u}$, and s_u^2 (the sample variance). Construct a graph of both the *true* and the *estimated* regression lines.

Input. Test your program with the following sets of input data.

http

Data Set 1	*Data Set 2*
$n = 50$ and 200	$n = 50$ and 200
$X_i = 0,\ i = 1, \ldots, n/2$	$X_i = 1,\ i = 1, \ldots, n/2$
$X_i = 1,\ i = n/2 + 1, \ldots, 100$	$X_i = 3,\ i = n/2 + 1, \ldots, 100$
$\alpha = 5, \beta = 1$	$\alpha = 4, \beta = 1.5$

Output. Output $\bar{y}$, $\hat{\alpha}$, $\hat{\beta}$, $\bar{u}$, and s_u^2 along with 50 of the sample points Y_i for purposes of drawing a scatter diagram.

Discussion and Questions. In the Introduction of Chapter 7, we described the properties of the random variable u, namely that u has some specific probability distribution with expected value 0 and variance σ_u^2. Also, we assumed that values of u are independent of each other in repeated sampling. These assumptions imply that the expected value of Y (the mean of Y) is $\alpha + \beta X$, since X is a known constant. Another way of stating this problem is that the mean value of Y for a given value of X is a linear function of X, with specific values of Y oscillating about the line in repeated sampling by a random error, u. This result is precisely what our computer experiment is performing. Here, the u_i is a random sample from a normal distribution with mean 0 and variance 1 (a common assumption in linear regression analysis), the X_i are constants specified by the input, and the Y_i is a random variable, since it is a function of the random variable u_i.

1. What is the estimated value of the mean of Y for each value of X? How does this value compare with the theoretical value?
2. What is the estimated value of α? How does this value compare with α?
3. What is the estimated value of α? How does this value compare with β?
4. What are the estimated values of E(u) and σ_u^2? How do these values compare with E(u) and σ_u^2?
5. Show that $\bar{y}$, $\hat{\alpha}$, $\hat{\beta}$, $\bar{u}$, and s_u^2 approaches $\alpha + \beta X$, α, β, 0 and 1, respectively, as the sample size increases.
6. Plot 50 of the observations on graph paper and draw in both the *true* and *estimated* regression lines.

Exercise 7.2

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations (u_i) from a population which is N(0,1). Let us pretend that the slope of line β is unknown. Using the 5% significance level, test the hypothesis that the slope of the line β is equal to β_0 (a constant). This is a two-tailed test.

Input. Test your program with the following sets of input values.

Data Set 1	*Data Set 2*
$\beta_0 = 0,\ \beta = 0$	$\beta_0 = 0,\ \beta = 1$
$n = 10$	$n = 10$
$x_i = 1,\ i = 1, \ldots, n/2$	$x_i = 1,\ i = 1, \ldots, n/2$
$x_i = 2,\ i = n/2 + 1, \ldots, 100$	$x_i = 2,\ i = n/2 + 1, \ldots, 100$
$\alpha = 5$	$\alpha = 5$

http

Data Set 3

$\beta_0 = 0, \quad \beta = 3$
$n = 10$
$x_i = 1, \quad i = 1,..., n/2$
$x_i = 2, \quad i = n/2 + 1,..., 100$
$\alpha = 5$

Output. Your program should output the T statistic for each data set.

Discussion and Questions. In this exercise, we are going to use the random variable T to test the hypothesis $\beta = 0$. Using these three data sets, it is possible to determine whether a decision concerning $\beta = 0$ is right or wrong. Therefore, it is possible to determine whether the decision is correct, a Type-I error has been made, or a Type-II error has been made. The principle purpose of this exercise is to determine whether the data presents sufficient evidence to indicate that X and Y are linearly related over the region of observation.

1. From each of the data sets, indicate whether there is insufficient evidence to reject the hypothesis that $\beta = 0$.
2. From each of the data sets, state whether your conclusion is correct, a Type-I error has been made, or a Type-II error has been made. Also state whether a certain type of error is possible for each data set.
3. If a hypothesis is already rejected by a sample of n = 10 observations, is it likely to be rejected by a sample of n = 100 observations? Why?
4. What would be the effect of changing the significance level and increasing the sample size?
5. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?
6. What is the consequence of using zero percent significance level? What does this mean in terms of Type-I and Type-II error?

Exercise 7.3

Statement of the Problem. Prepare and execute a program that will take a random sample of n observations (u_i) from a population which is N(0,1). Using the 5% confidence coefficient, compute a confidence interval for β, the slope of the line. Repeat the above experiment N times. Tabulate the number of times the slope of the line $\beta = \beta_0$ falls within the calculated interval.

Input. Test your program with the following sets of input values.

Data Set 1

$n = 10, \quad N = 100$
$x_i = 0, \quad i = 1, ..., n/2$
$x_i = 1, \quad i = n/2 + 1, ..., 100$
$\alpha = 5, \quad \beta_0 = 1$

Data Set 2

$n = 40, \quad N = 100$
$x_i = 0, \quad i = 1, ..., n/2$
$x_i = 1, \quad i = n/2 + 1, ..., 100$
$\alpha = 5, \quad \beta_0 = 1$

http

Output. Your program should output the number of times the slope of the line β falls within the calculated intervals, the sample slope $\hat{\beta}$ and the sample variance of the slope for each of the N experiments, and the total sample slope and variance of the slope.

Discussion and Questions. In this exercise, we are concerned with studying the relationship between X and Y. We shall pose a question here: once we have decided that X and Y are linearly related, what else can we do in examining the relationship in detail? Since we have estimated the slope of the line, how much confidence can be placed in the estimate $\hat{\beta}$? In order to confirm our confidence, we shall establish a confidence interval for β. It must be pointed out that we did *not* need any assumption on the distribution of the random variable u in making a point estimate of β. Once we are interested in making a probability statement concerning β, then we must assume we know the exact distribution of ⊔. In this exercise, we assume that the random variable u follows a normal distribution with mean 0 and variance 1.

1. For each data set, how does the tabulated number of times the slope β falls within the calculated interval compare with the theoretical result, that is, N × .95?
2. What is the consequence of increasing the sample size without changing the confidence coefficient?
3. Why do the lengths of the confidence intervals of the slope β change from sample to sample even when the sample size remains the same?
4. Regardless of the sample size, one can obtain a confidence interval. Why does one prefer a large sample? Comment about your point estimates of $\hat{\beta}$ and s_u^2.
5. Is it necessary to know in this exercise that we are sampling from a normal distribution? Why?
6. What is the consequence of using zero percent confidence coefficient?

Exercise 7.4

Statement of the Problem. This exercise is concerned with testing a hypothesis for the *expected* value of Y for a given value of X. Perform this exercise exactly like Exercise 7.2, except test the hypothesis that the expected value of Y for a given value of $X = x_0$, $E(Y \mid X = x_0) = \alpha + \beta x_0$ is equal to E_0, a constant. Use the data sets and answer all pertinent questions described in Exercise 7.2.

Exercise 7.5

Statement of the Problem. This exercise is concerned with establishing a confidence interval for the *expected* value of Y for a given value of X. Perform this exercise exactly as you did Exercise 7.3, except establish a confidence interval for $E(Y \mid X = x_0) = \alpha + \beta x_0$. Use the data sets and answer all pertinent questions described in Exercise 7.3.

http

Exercise 7.6

Statement of the Problem. This exercise is concerned with establishing a confidence interval for a particular value of Y for a given value of X. Perform this exercise exactly as you did Exercise 7.3, except calculate confidence intervals for Y values.

http

8

Analysis of Variance

Prerequisites

The exercises of this chapter require a basic understanding of analysis of variance.

Introduction

In most experiments that are to yield a valid conclusion, several variables need to be controlled aside from the one or more being investigated. In some cases, the interfering variables can be controlled by laboratory techniques; in others, such control may be possible only through statistical design.

One of the most useful techniques for increasing the sensitivity of an experiment is the designing of the experiment in such a way that the total variation of the variable being studied can be separated into components that are of experimental interest. The splitting of the total variation in this manner enables the experimenter to utilize statistical methods to eliminate the effects of certain interfering variables and hence increase the sensitivity of his experiment. The *analysis of variance* is a technique for carrying out the analysis of an experiment designed from this point of view.

Before the equations which define the algorithm used in the analysis of variance are given, the technique will be studied by a simple example. Our data consists of two

http

independent random samples of size n. Two subscripts will be used to completely identify any observed value and the observations will be denoted by Y_{ij}. The first subscript, i, will denote the sample and the second subscript, j, will denote the individual observation within the sample. Thus Y_{14} is the fourth observed value in the first sample. The term *treatment* will be used synonymously with the term *sample;* that is, an individual observed value from sample one is equivalent to an observed value from treatment 1.

For data of the type described above, we will assume the mathematical model

$$(1) \qquad \begin{cases} Y_{ij} = \mu_1 + \beta x_i + \epsilon_{ij}; \quad i = 1, 2, \; j = 1, 2, \ldots, n \\ \{e_{ij}\} \text{ are independently } N(0, \sigma^2), \end{cases}$$

where $\beta = \mu_2 - \mu_1$ is the (unknown) difference in the two treatments and is the object of interest in the experiment. The *dummy variable*, x_i, is defined as

$x_1 = 0$ if the individual observed value Y_{ij} is from sample 1;

$x_2 = 1$ if the individual observed value Y_{ij} is from sample 2.

Thus, when we are predicting the variable Y_{1j}, the dummy variable x_1 is set equal to zero and the model is

$$Y_{1j} = \mu_1 + \epsilon_{1j}; \quad j = 1, 2, \ldots, n$$

For treatment 2, x_2 is set equal to one, and the model is

$$Y_{2j} = \mu_1 + \beta + \epsilon_{2j}$$

or

$$Y_{2j} = \mu_2 + \epsilon_{2j}$$

since $\beta = \mu_2 - \mu_1$. It should be pointed out that the mathematical model (1) is exactly of the form described in Chapter 7.

Another mathematical model for the data of the type under consideration is

$$(2) \qquad \begin{cases} Y_{ij} = \mu + \Upsilon_i + \epsilon_{ij}; \quad i = 1, 2, \; j = 1, 2, \ldots, n \\ \{e_{ij}\} \text{ are independently } N(0, \sigma^2), \end{cases}$$

which states that any observed value Y_{ij} is equal to the overall mean μ for all the populations, plus the deviation Υ_i of the i-th population mean μ_i from μ, plus a random deviation ϵ_{ij}. In other words,

http

$$\mu = (\mu_1 + \mu_2)/2$$
$$\Upsilon_i = \mu_i - \mu$$
$$\epsilon_{ij} = Y_{ij} - \mu_i = Y_{ij} - \mu - \Upsilon_i.$$

The Υ_i are known as *treatment effects*. For this model as well as model (1), we assume that μ is an unknown parameter and Υ_i are unknown parameters. It is left as an exercise to show that model (1) is equivalent to model (2).

Let us now return to our discussion of model (1). As in the regression model of Chapter 7, we wish to estimate μ_1, β, and σ^2 statistically on the basis of our sample observations on Y. We may also wish to test hypotheses or establish confidence intervals about these parameters. In order to accomplish this, we estimate the true line $Y = \mu_1 + \beta x$ by $\hat{Y} = \hat{\mu}_1 + \hat{\beta}x$, where

$$\hat{\beta} = \frac{\sum_{i=1}^{2}\sum_{j=1}^{n} X_i (Y_{ij} - \overline{Y})}{\sum_{i=1}^{2}\sum_{j=1}^{n} (X_i - \overline{X})^2},$$
$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}.$$

It should be clear that $\overline{Y} = (\overline{Y}_1 + \overline{Y}_2)/2$ and $\overline{X} = 1/2$. After some simple algebra, it can be shown that $\hat{\beta} = \overline{Y}_2 - \overline{Y}_1$ and $\hat{\alpha} = \overline{Y}_1$. This proof is left as an exercise to the student.

The following exercises are concerned with the above model (1) and the problem of making statistical inferences concerning the model.

Exercise 8.1

Statement of the Problem. Prepare and execute a program that will take two independent random samples, each of n observations (ϵ_{ij}) from a population which is N(0,1). This exercise utilizes model (1) from the introduction. From these observations, find $\hat{\mu}_1$, $\hat{\beta}$, $\overline{\epsilon}$ and s_ϵ^2 (the estimate of σ^2). Construct a graph of both the *true* and *estimated* regression line.

Input. Test your program with the following sets of input data.

Data Set 1	*Data Set 2*
$n = 50$	$n = 50$
$\mu_1 = 1$	$\mu_1 = 2$
$\beta = 2$	$\beta = 0$

http

Output. Output $\hat{\mu}_1$, $\hat{\beta}$, $\bar{\epsilon}$ and s_ϵ^2 along with the sample points Y_{ij} for the purpose of drawing a scatter diagram.

Discussion and Questions. In the Introduction of Chapter 8, we described the properties of the random variable ϵ, namely that ϵ has some specific probability distribution with expected value 0 and variance σ^2. Also, we assumed that values of ϵ are independent of each other in repeated sampling. These assumptions imply that the expected value of Y (the mean of Y) is $\mu_1 + \beta X$, since X is a known constant. Another way of stating this problem is that the mean value of Y for a given value of X is a linear function of X, with specific values of Y oscillating about the line in repeated sampling by a random error, ϵ. This result is precisely what our computer experiment is performing. Here, ϵ_{ij} are a random sample from a normal distribution with mean 0 and variance 1. The X_i are constants specified by the input and the Y_{ij} is a random variable, since it is a function of the random variable ϵ_{ij}. The student is advised to reread Exercise 7.1

1. What is the estimated value of the mean of Y for each value of X? How does this value compare with the theoretical value?
2. What is the estimated value of μ_1? How does this value compare with μ_1?
3. What is the estimated value of β? How does this value compare with β?
4. What are the estimated values of $E(\epsilon)$ and σ^2? How do these values compare with $E(\epsilon)$ and σ^2?
5. Show that $\hat{\mu}_1$, $\hat{\beta}$, ϵ and s_ϵ^2 approaches μ_1, β, 0 and 1, respectively, as the sample size increases.
6. Plot 50 of the observations on graph paper and draw in both the *true* and *estimated* regression lines.

Exercise 8.2

Statement of the Problem. Prepare and execute a program that will take two independent random samples, each of n observations (ϵ_{ij}) from a population which is N(0,1). This exercise utilizes model (1) from the Introduction. Using the 5% significance level, test the hypothesis that the slope β is equal to β_0 (a constant). This is a two-tailed test.

Input. Test your program with the following sets of input values.

Data Set 1	*Data Set 2*	*Data Set 3*
$\beta_0 = 0$, $\beta = 0$	$\beta_0 = 0$, $\beta = 1$	$\beta_0 = 0$, $\beta = 3$
$n = 10$	$n = 10$	$n = 10$
$\alpha = 1$	$\alpha = 1$	$\alpha = 1$

Output. Your program should output the T statistic for each data set.

Discussion and Questions. In this exercise, we are going to use the random variable T to test the hypothesis $\beta = 0$. Using these three data sets, it is possible to determine

http

whether a conclusion concerning $\beta = 0$ is right or wrong. Therefore, it is possible to determine whether the conclusion is correct, a Type-I error has been made, or a Type-II error has been made. The principle purpose of this exercise is to determine whether the data presents sufficient evidence to indicate that X and Y are linearly related over the region of observation, or in terms of analysis of variance, whether μ_1 is equal to μ_2.

1. From each of the data sets, indicate whether there is insufficient evidence to reject the hypothesis that $\beta = 0$.
2. From each of the data sets, state whether your conclusion is correct, a Type-I error has been made, or a Type-II error has been made. Also state whether a Type-I error or a Type-II error is possible for each data set.
3. If a hypothesis is already rejected by a sample of n = 10 observations, is it likely to be rejected by a sample of n = 100 observations? Why?
4. What would be the effect of changing the significance level and increasing the sample size?
5. Is it necessary to know in this exercise that we are sampling from a random normal population? Why?
6. What is the consequence of using zero-percent significance level? What does this mean in terms of Type-I and Type-II error?

http

Appendix I
Basic Elements of FORTRAN IV

Introduction

The small portion of the FORTRAN IV language which is presented here is sufficient to write programs that will aid you in completing this course. Your competence will increase as you continue to practice this skill. Although much of FORTRAN will not be explained here, you will be able to resolve the additional features of the language by referring to the references which are included.

The FORTRAN IV Language

A *digital computer* is a machine which can perform arithmetic on numbers; that is, it can add, subtract, multiply, and divide. Several features set it apart from an ordinary calculating machine. Chief among them is its ability to store vast volumes of numbers in its "memory," together with a list of operations to perform on those numbers. This list of operations is called a program. Each operation in the list is called an order or an instruction. Since a computer can cope only with numbers, the instructions are coded to appear as numbers too. As long as the program is stored in the memory of the computer (which is the same place that data is stored) the computer can also operate upon the program so as to modify the program as the

http

computation proceeds. Thus, a short program can present enough control information to a computer to keep it running for a long time, even without end. Every computer is composed of five parts: input unit, memory unit, arithmetic unit, control unit, and output unit. Data and programs to be inputted to the computer must be punched into cards. A card is composed of eighty columns, and each column is used to represent one character. Every student should sit down at a keypunch and punch every character just to see what the result is on the card.

A deck of cards is read by the input section of the computer and placed in the memory. To use a number which is stored in memory, the computer calls it by an address (like a house number). Stored in memory are data and programs. The program is executed one number after another by the control section of the computer, with most of the actual computation carried on in the arithmetic section. Data which is ready to be outputted is directed to the output section by appropriate orders, there to be printed on a page by a line printer. It should be emphasized that all movement of data, be it input, computation, or output, must be by appropriate instructions in memory.

The *characters* with which FORTRAN IV is written are:

a. 26 upper case letters: A B C ... N Ø P ... X Y Z
b. 10 digits: 0 1 2 3 4 5 6 7 8 9
c. 11 special characters: + − * / = ( ) . , '
d. a blank (sometimes denoted by ␣)

Various combinations of these characters are used to form *names* of items such as *variables* and *constants*, to write *expressions* involving variables, constants, or other quantities, and to make up the *statements* of the language. It must be emphasized that every piece of data in the machine must be assigned a name, which acts as the address of that piece of data in memory. A particular name may stand for only one thing in a Fortran program.

The *rules for names* in FORTRAN IV are:

a. A name is composed of at least one and not more than six characters.
b. The first character must be a letter.
c. The only characters used must be letters or digits (no special characters such as +, −, etc.).

Examples of *acceptable* names:

X I AS MAX AIB2C3

Examples of *unacceptable* names:

3X 2I MAX.1 ABCDFGD WILLIAM

http

Variables and *constants* may be of five different *types: integer* (which is often referred to as *fixed-point), real* (which is often referred to as *floating point), double-precision, complex,* and *logical.* Only the integer and real types will be considered in this discussion.

The *rules for integer variable names* in FORTRAN IV are:

a. The name usually starts with I, J, K, L, M, or N.

b. One of these letters is sufficient to indicate the name for an integer variable.

Examples of integer variable names:

IX J12 IMAX K ITEST

The *rules for integer constants* in FORTRAN IV are:

a. An integer constant may have from 1 to 11 digits, and a sign (+ or −) if necessary. (Actually the integer must be less than 2^n, depending on the computer; n = 35 on the GE635 and UNIVAC 1108; n = 31 on the IBM 360; n = 59 on the CDC 6000; n = 15 on the IBM 1130, etc.)

b. An integer must *not* have a decimal point in it.

Examples of integer constants:

0 1 −17 5286 +017 −1

The *rules for real variable names* in FORTRAN IV are:

a. The name usually begins with any of the letters A to H, or Ø to Z.

b. One of these letters is sufficient to indicate the name for a real variable.

Examples of real variable names:

AS VENGTH XK W123 B

The *rules for real constants* in FORTRAN IV are:

a. A real constant must have a decimal point.

b. It must have at least 1 digit with a sign if necessary. (The maximum number of digits depends on the computer)

c. It may be followed by the letter E (for Exponent of 10) which *must* be followed by a one or two digit signed or unsigned number.

Examples of real constants:

1. .0001 2.3 −573.216 123456789.
2.74E3 (means 2.74 x 10^3 = 2740.)
2.7E − 3 (means 2.7 x 10^{-3} = .0027)
100E5 (means 100. x 10^5 = 10000000.)

http

An *arithmetic expression* consists of arithmetic constants, arithmetic variables and arithmetic function calls (discussed later in these notes) which are combined by arithmetic operators and parentheses. The *arithmetic operators* are as follows:

Operator	Operation	Algebraic Expression	FORTRAN Expression
+	addition	a + b	A+B
−	negation	−a	−A
−	subtraction	a − b	A−B
*	multiplication	ab, or a · b	A*B
/	division	a/b, or $\frac{a}{b}$	A/B
**	exponentation	a^b	A**B

Here the first "−" is a "unary" operator which is applied to only one operand, whereas the second "−" is a "binary" operator, since it combines two operands. The "+" operation may also be applied as a unary operator, although it causes no action. The operands A and B must be of the same type, integer or real, except in the case of exponentation. In that case, a real base A may be raised to an integer power N (for example, X**2 is allowed as well as N**2 and X**2.47). Division of an integer causes truncation (that is, the fraction is lost). For example, 3/2 equals 1, not 1.5.

An arithmetic expression is defined as one of the following:

a. an arithmetic constant (integer or real type),

b. an arithmetic variable (integer or real type),

c. an arithmetic function call (see definition later),

d. an arithmetic expression enclosed in parentheses,

e. an arithmetic expression preceded by + or −,

f. any two arithmetic expressions connected by a binary arithmetic operator.

Three *additional rules* must be followed; first, no two operators may occur in sequence; second, no operation symbol may be assumed to be present (for example, (A)(B) is invalid and must be written as (A)*(B);) and third, the expression A**B**C is invalid and must be written as (A**B)**C or as A**(B**C).

Since operations inside parentheses are performed before those outside, the order of performance of the calculations can be completely specified by parentheses. The *rules of precedence* for arithmetic operations allow parentheses to be omitted in many cases. These rules state that the operation listed first or highest in the following list will be performed first:

Function evaluation
** and unary + or −
* and /
+ and −

with the further provision that of any two or more operators at the same level in this list the leftmost one in the expression will be performed first. Examples of such expressions are:

−5	X**2	2/4+6/N
2.754	A+B*C	4.0/(X−3.0)+7.65
A	X+1.0	(A+B)*(C−D)
I	N+3	(E**X−E**(−X))/2.0

The rules of precedence cause the following expressions to be equivalent:

−A*B−C/D−E and (((−A)*B)−(C/D))−E
A−B/C*D/E and A−((B/C)*(D/E))

Finally, expressions are made into statements (which really correspond to an instruction in the computer) by placing a variable name on the left side of an equal sign, then an equal sign, followed by an expression.

An *arithmetic assignment statement* is of the form

v = e,

where

v is a variable (subscripted or not)
e is an arithmetic expression.

This statement causes the expression on the right of the equals to be evaluated, and the value obtained is then assigned to the variable on the left. This statement *must not be* interpreted as an equation in the algebraic sense, where it would be immaterial whether one wrote v = e or e = v. Although the variable and the expression are usually of the same type—integer or real—this is not necessary. The value of a real expression will be truncated (cut off to the next lower integer) and converted to an integer before being assigned to an integer variable. Similarly, the value of an integer expression will be simply converted to real before being assigned to a real variable. This conversion of type is useful in many situations. Some examples of assignment statements are

N = 5	X = 2.4729
N = N+1	X = Y+3.47
N = I+I	X = (R−S/2.0)**0.5
N = X+1.0	X = N/2

Before a variable may be used in an expression, it must be given a value. This assignment may have been made by giving it a value in a previous assignment

statement where it occurred on the left side. Often it is desirable to read or *input* the values of certain variables as *data*. Certainly, after completing the computations for a problem, it is desirable to have the computer write or *output* the results for our use. There are a number of ways to input and output values in FORTRAN, but we shall introduce only the most simple ones.

The general form of *input and output statements* is:

READ(5,n1) *list1*
WRITE(6,n2) *list2*

where n1 and n2 are the numbers (integer constants) of *format statements*, and list1 and list2 are sequences of one or more *names* separated by commas. The names are the names of variables, arrays, components of arrays, or groups of components. Also, list1 and list2 may be empty.

Some examples of input and output statements are

READ(5,17)K,X,Y,Z
WRITE(6,18)V,W,N

The integer constant 5 in the above READ statement designates that the data is to be read from the cards furnished by the user in the deck following his program. This deck is usually called the *data deck*. The integer constant 6 in the above WRITE statement designates that the output is to be printed on the high speed printer as part of the listing normally returned to the user.

Every input or output statement of the general form described above must have a *format statement* associated with it. The general form of a format statement is

n FORMAT(S1,S2, ... ,Sm)

where n is the integer constant referred to by an input or output statement and the S's are specifications or groups of specifications. The *specifications* to be discussed here are of the following form:

Iw Fw.d Ew.d wX wH

(only wH is followed immediately by w characters), where w and d designate an unsigned integer. Each specification must be followed by a comma, except that commas after X and H specifications are optional.

The following are examples of how quantities appear according to the indicated specification. Each of the quantities appearing on the left is printed as shown on the right according to the indicated specification. Each of the quantities appearing on the right is inputted as shown on the left, in parentheses, according to the indicated specification. The symbol ⊬ denotes a blank.

http

Internal value	Specifications	External appearance
21 (21)	I5	␣␣␣21
−21 (−21)	I5	␣␣−21
21 (21)	I3	␣21
7.6247 (8.0)	F4.0	␣␣8.
7.6247 (7.6)	F5.1	␣␣7.6
7.6247 (8.0)	E7.0	0.8Eb01
7.6247 (7.6)	E9.2	␣0.76E␣01
−7.6247 (−7.6)	E9.2	−0.76E␣01
–	4X	␣␣␣␣
–	3HWGB	WGB

The following are safe rules of E and F:

a. In Fw.d, w must exceed d + 2.

b. In Ew.d, w must exceed d + 6.

These rules can be violated, but not until you become expert.

Groups of specifications are formed by enclosing lists of individual or group specifications in parentheses. Any individual or group specification may be preceded by a repetition number, except for the individual specification wX and wH . . . . Here is an example of the preceding rule:

```
FORMAT(I2, F5.2, I2, F5.2, I6, I6, 3HWGB, 3X, F6.1)
FORMAT(2(I2, F5.2), 2I6, 3HWGB, 3X, F6.1)
```


The preceding two FORMAT statements are equivalent.

Let us return briefly to the discussion of variables, in order to define a vector in FORTRAN. A *vector* is a list of variables, all of the same type, which has a single name. The individual elements or variables of a vector are referred to by means of subscripts applied to its name. The *dimension* (number of *elements*) in the vector must be specified, and this identification is done with a statement of the form:

```
DIMENSION A(15), M(10)
```

which means that the real vector A has fifteen elements A(1), A(2), ..., A(15), and the integer vector M has ten elements M(1), M(2), ..., M(10).

A subscript is a simple arithmetic expression which may take on any of the following forms where v stands for an unsigned nonsubscripted integer variable and c and d represent any unsigned integer constants:

```
v    c    v+d    v−d    c*v    c*v+d    c*v−d
```

The value of the subscript expression cannot be zero, negative, or greater than the dimension (number of elements) of the vector which is referenced. Examples of *subscripted variables* are:

http

A(13) M(N) A(I+2) M(2*I) A(3*N+4) TAG(2*K−4)

The *STOP* statement brings the execution of the program to a halt, transfers control back to the "system," which then clears the program out of the computer and puts it to work on another job. Several such statements may appear in a Fortran program, but only one will be executed for any particular running of the program.

The *END* statement must be the very last statement of any Fortran program, and so only one such statement may appear. This statement simply informs the Fortran compiler that there are no more instructions to follow

The *execution* of statements in a Fortran program is normally carried out in the sequence in which they appear, though this sequence can be altered in several ways. The simplest *transfer of control* is a statement which tells the computer to "go to" another statement and execute it next. This statement requires that one be able to refer to other statements. To fulfill this requirement, any statement may be numbered for reference elsewhere. One special statement which does absolutely nothing but allow a point in a program to be numbered for such reference is the continue statement:

```
10    CØNTINUE
```

The *go to* or unconditional transfer statement is:

80

```
GØ TØ 10
```

where 10 may be replaced by any appropriate statement number which is on the statement to be executed next. The "next" statement may occur before the GØ TØ statement in the program, in which case a cyclic sequence or "loop" is established. For example, the following sequence of two statements is such a loop; it does absolutely nothing useful and yet uses up computer time.

```
10    CØNTINUE
      GØ TØ 10
```

Such "infinite" loops should be avoided whenever possible.

The simplest decision making in Fortran involves executing another statement (or not), depending on the truth or falsity of a *logical* expression. The simplest of these logical expressions is a relation. A relation consists of two arithmetic expressions of the same type—integer or real—joined by one of the following relational operators:

http

Operator	Relation	Example	Value
.EQ.	equals	1.+1. .EQ. 2.	T
.NE.	not equal	N .NE. N	F
.LT.	less than	X .LT. X+1.	T
.GT.	greater than	N .GT. N	F
.LE.	less or equal	O. .LE. X**2	T
.GE.	greater or equal	N**2 .GE. N	T

More complex logical expressions are not included in this description.

The *logical if statement* consists of the word IF followed by a logical expression in parentheses followed by a simple statement such as a GØ TØ, a READ, a WRITE, a STØP, or an assignment statement. As examples of these, the following statements are valid:

```
IF(X.GE.0.)GØ TØ10
IF(A.EQ.0.)STØP
IF(N.NE.10)X = N + 10
```

In each case the simple statement will be executed only if the relation holds true; otherwise, the next statement in sequence will be executed. By using combinations of these decision-making statements, more complex decisions can be indicated in programs.

The *DO statement* makes it possible to execute or "do" a set of statements some number of times with an *index* variable being changed each time they are executed. The last statement in the set is numbered and this number included in the DO statement. The first value and terminal value of the index must be given. The amount the index is to be stepped or changed after each pass through the instructions can be given; otherwise it will be assumed to be 1. The statement has the form

```
DO n i = f,t,s
.
.
.
n  (last statement to be executed)
```

where

- n is the number of the last statement to be executed,
- i is the nonsubscripted integer index variable,
- f is the first value of i,
- t is the terminal value of i,
- s is the step size or amount which is added to i after each pass and may be omitted if a step size of 1 is desired,

f,t,s are the DO parameters, which must be positive integer constants or integer variables that have positive values with t greater than f.

Examples of the use of the DO statement are

```
      S = 0.0
      DO 10 I = 1,15
10    S = S + A(I)
```

which sets S to the sum of the first 15 elements of A, and

```
      DO 20 N = 1,100,2
      A(N) = A(N) + 1.0
20    A(N+1) = A(N+1) — 1.0
```

which adds one to the odd subscripted elements of A and subtracts one from the even ones for the first hundred such elements.

Many of the standard mathematical functions are available in Fortran IV. A few of these are listed in the table below. The type of the argument and value must be strictly followed. The "arithmetic function call," as shown under "Example" below, is included in an arithmetic expression as described earlier.

Function	Name	Argument	Value	Example	Meaning
Real Absolute Value	ABS	Real	Real	ABS(X)	| x |
Integer Absolute Value	IABS	Integer	Integer	IABS(I)	| i |
Convert Integer to Real	FLOAT	Integer	Real	FLOAT(N)	
Convert Real to Integer	IFIX	Real	Integer	IFIX(S)	
Exponential	EXP	Real	Real	EXP(X)	e^x
Nautral Logarithm	ALØG	Real	Real	ALØG(X)	ln x
Common Logarithm	ALØG10	Real	Real	ALØG10(X)	log x
Trigonometric Sine	SIN	Real	Real	SIN(X)	sin x
Trigonometric Cosine	CØS	Real	Real	CØS(X)	cos x
Square Root	SQRT	Real	Real	SQRT(X)	$\sqrt{x}$

Examples of the use of these functions follow:

```
Y = ABS(X) + 1.0      Y = ALØG(X)**2
M = IABS(I) + 1       Y = SIN(X) + CØS(X)
N = IFIX(X) + 1       Y = SQRT(X + 2.0)
```

A Fortran program consists of a sequence of statements as described above. Each statement is normally punched on a single card. The statement must be punched in columns 7 through 72. The statement number—if any—is punched in columns 2 through 5. Column 6 is used to indicate "continuation" cards for statements which will not fit on one card. Columns 73—80 may be used for program identification and sequential numbering of cards or other purposes.

http

Some of the statements of a Fortran program must precede others. For simple programs, the following ordering of the statements is usually used: first, the *type* (not discussed here) and *dimension* statements; second, the *namelist* (not discussed here) statements; third, the operational statements, (*assignments, stop* statements, *if* statements, *do* statements, etc); and finally the *END* statement.

Explanatory notes or "comments" may be inserted anywhere in the program by putting the letter "C" in the first column of a card. Anything can then be punched in these cards. These comments may be used both for your own information and to explain the program to others. In particular, it is useful to put a title on the program with a comment card which appears as the first card.

Since blanks are usually ignored in Fortran programs, they can and should be used to make the program as readable as possible for the people involved.

http

Appendix II
Generating Pseudorandom Numbers, U(0,1)

In this appendix we are primarily concerned with methods of generating random numbers with digital computers, leaving the matter of applications to Chapters 2–8.

The method generally used to generate pseudorandom numbers on a digital computer is a congruential method—namely,

$$X_{i+1} = (\lambda x_i + c) \quad \text{(modulo P)}.$$

The above equation in its present form is called a mixed congruential method. If $\lambda = 1$, then the method is called additive congruential; if $c = 0$, then the method is called multiplicative congruential. As pointed out by Coveyou and MacPherson[1], mixed and multiplicative methods are better than additive methods. Furthermore, since they found very little difference between mixed and multiplicative methods, the latter will be described here; it offers advantages in terms of computational speeds and space requirements.

Given an initial starting value x_0, a constant multiplier λ, then $x_i + 1 = \lambda x_i$ (modulo P) yields a multiplicative relationship (modulo P) for any value of i over the

[1] Coveyou, R. R., and R. D. MacPherson, "Fourier Analysis of Uniform Random Number Generators," *J.ACM*, XIV(1967), 100–119.

sequence $\{x_1, x_2, ..., x_1 ...\}$. Hence, we have formed a sequence of integer numbers $\{x_1, x_2, ..., x_i, ...\}$. We also have that $x_i < P$ for all x_i. From the sequence of integers $\{x_i\}$, rational numbers in the unit interval (0,1) can be obtained by forming the sequence $\{r_i\} = \{x_i/P\}$. Our computer program to generate random numbers in the unit interval will employ a modulus $P = 2^m$, where m represents the word size of the binary computer in bits. The following are examples of P for commonly used computers.

$P = 2^{31}$	$= 214783648$	for IBM 360
$P = 2^{35}$	$= 34359738368$	for GE 600 or UNIVAC 1108
$P = 2^{59}$	$= 2^{35} \cdot (16777216)$	for CDC 6000

Since we desire to write a program to generate $\{r_i\} = \{x_i/P\}$, we shall write $r_i = X_i \cdot 2^{-m}$. As an example, for m = 31,

$$r_i = X_i \cdot (.4656613E - 9)$$

or for m = 35,

$$r_i = X_i \cdot (.29103805E - 10).$$

The multiplicative procedure for generating random numbers on an m-bit binary machine may be summarized as follows:

1. Choose any odd number x_0 (integer) as a starting value which is less than $P = 2^m$.
2. Choose an integer λ, which is the constant multiplier. (As suggested in the reference of Footnote 1, $\lambda = 5^{13} = 1220703125$ is a good selection.)
3. Compute λx_0 using integer arithmetic. This product consists of 2m bits of which the high order m bits are discarded and the low order m bits represent x_1. (This discarding of bits in integer arithmetic is done automatically in FORTRAN.)
4. Calculate $r_1 = x_1 \cdot 2^{-m}$ to obtain a pseudorandom number on the unit interval.
5. Each successive random number x_{i+1} is obtained from the low order bits of the product λx_i.

The following is a FORTRAN subroutine utilizing the preceding algorithm with m = 35.

```
FUNCTION RCM(IST)
IST  = IST*1220703125
RCM = IST
RCM = RCM*0.291038305E−10
RETURN
END
```


http

Examples of other values of λ are given by Coveyou and MacPherson[2]. The output RCM is a floating-point random number between 0 and 1. The main program that calls this generator must provide an initial odd integer value of IST, which must be less than P. The function RCM(IST) can be called in an expression in any place where a variable is permitted.

The exercises of this book use the expression "prepare and execute a program that will take n observations from a population which is U(0,1)." This statement is equivalent to issuing the above FORTRAN CALL n times.

The table of random digits at the end of this section illustrates a set of 1000 random digits, that is, a 1000 values of x_i. The digits have been arranged in columns of five, and the columns and rows have been further divided into groups for ease of reading. If it is decided to generate a series of uniformly distributed random numbers between 0 and 1 (this scheme will be denoted by U(0,1)) to an accuracy of 5 decimal digits, then the first column on the left can be read, and when the column is finished the next column can be used. Using this method, the first five random numbers from U(0,1) are

.26169
.41927
.69241
.13559
.68550

If numbers with more than five digits are needed, columns are combined. It is not necessary to start with the first column. In fact, for proper use, particularly when a calculation is being repeated with different sets of random numbers, the starting points should be chosen at random. This selection may be made by using a random number to decide upon the page, column, and row at which to start. It is possible to read in the table of random digits at the end of this section, but the number of random digits may be large and it is somewhat inconvenient to prepare data in this form.

[2] *Ibid.*

http

26169	36649	07723	75898	48330	13207	54974	30717	43680	73909	48752	14447
41927	33486	81536	17992	22643	74537	37447	85845	92848	99579	55842	55822
69241	13730	56976	88393	51771	90337	35182	81702	62801	09418	12556	05880
13559	96241	33606	27757	47101	24901	68843	34428	94511	89226	82765	02189
68550	80862	09221	95752	05476	51882	64621	05398	61251	76754	14945	22659
27855	66026	16043	80067	78932	21557	78174	59231	71469	60652	05268	95937
73756	73923	06871	04981	49307	19255	80243	68828	22003	91680	91582	73681
34143	46533	88837	14054	96633	90722	64086	24450	18038	94456	70934	34222
86711	13272	89923	79233	67708	47718	22395	41392	63726	05516	80799	24659
76768	68564	30900	13628	94745	04616	28860	98267	14857	00983	83662	85912
02578	35097	62846	59115	73633	34238	47018	75434	72511	14609	26117	42759
09642	26062	17890	41268	05312	69062	76558	00610	84066	84483	47131	78406
22814	57139	08353	23662	99345	10025	34734	46184	05497	04184	95641	50599
32670	96527	96925	31067	96921	45459	58039	60276	66357	91649	05127	43825
40033	78789	48048	48574	15092	16691	96801	51070	66349	26563	60329	89629
50051	45796	58159	41177	31018	07131	62855	81440	19899	44521	78754	24576
94750	15559	49940	75846	68638	67788	77947	68418	52628	13767	17138	92887
50454	17208	03244	65623	09790	30045	58024	19528	41140	57779	00514	38294
03988	67932	82843	92769	46789	93963	56074	75510	75024	24486	59041	67124
09067	05133	10660	05495	81053	17913	73666	66477	07480	40397	45263	02162
61771	24585	17642	20310	58493	72798	38843	26029	48474	25424	91925	84312
11526	12889	97921	99513	20999	61663	35532	45359	41827	81680	58046	16597
11484	99984	75438	10865	40218	42966	57128	86876	55149	07029	98357	82529
99713	95021	80659	13970	87761	33287	44393	14388	99018	21677	78802	08321
90111	06357	22570	55392	82899	02770	55352	34360	79314	33582	48229	76169
39432	83966	28580	92049	46997	14063	02312	80298	74119	05961	63946	32800
44352	60045	12507	99909	80888	93051	78690	09767	14071	27678	56282	75661
09962	01087	18294	62514	69138	57154	85784	21097	32606	24794	04827	36060
18608	22221	74617	71377	02730	64474	57153	34295	39969	89066	86493	16383
15480	52079	83028	07773	98773	84579	65763	18195	62413	36659	43059	75881
14847	73989	49814	12953	83944	29188	86547	83380	75494	47877	03352	97085
18364	06761	59201	91325	95554	18561	74550	97909	47856	76172	82729	02359
24337	25858	73078	28618	41371	96859	06300	20377	99431	52221	70002	27635
94367	63230	29886	44573	47384	47252	19567	82341	65436	46364	99478	97854
31267	61587	02844	37181	09798	95083	74173	89637	45103	90157	93259	86148
09670	96400	65140	13015	55179	24828	43997	49633	01515	95348	46460	76299
88228	16409	97278	35660	00163	74172	37841	68938	66486	46042	15508	85489
19805	90912	59209	61812	89629	05961	34309	03596	91865	77344	57170	41896
63581	89630	37419	72033	79748	96311	03351	81310	76193	91254	53471	49151
90458	58416	64617	40730	19809	14669	74646	52712	83096	98100	45156	07184
60693	91598	97197	27367	74230	09093	00451	66221	55341	09291	83854	96511
40143	39233	23113	08453	13656	47539	48102	98514	15953	44669	00598	70473
09051	25055	60330	07327	91607	03434	44801	10146	35289	55241	75863	37460
04756	88394	93512	16280	10387	11322	84860	33342	30498	99573	83766	52669
27252	00858	84093	43047	69000	55852	97050	36507	51249	49631	70601	76402
23996	46043	76398	51185	64633	94907	57215	47467	93154	39338	15351	93459
57862	88073	09961	01390	35326	55132	18801	54990	67775	06648	07338	25638
49669	92227	35705	38263	62655	75669	21469	45601	45636	17401	63672	06883
74159	98455	21130	04575	10853	13358	25433	70238	76130	96741	96651	28116
75864	86055	17176	02545	10278	60193	06696	72770	75748	31400	28762	96269
58742	05290	46910	59179	08621	61324	35817	49913	63517	43590	40452	89563

http

91002	89233	63696	90188	92771	59367	50394	49572	10063	74831	23319	28555
54023	96617	13999	94520	72746	28320	18893	52143	54201	55453	98874	19988
51764	73973	27495	49044	06590	97859	80997	16806	28468	73084	50542	67977
84588	62832	44620	10407	32594	06293	37126	12335	22488	90891	04044	84560
78888	15282	79235	68601	62941	57965	47275	89471	00595	03854	78835	69139
51436	18653	01548	34270	79835	98378	86844	89366	77598	58865	20730	63861
74850	67617	13962	79295	63374	57827	94600	20152	44120	01922	13395	29447
98093	59474	79998	87667	67364	51814	35438	71152	20394	53208	53855	23518
63410	25923	53944	11209	47040	64036	88092	71768	04938	48323	64661	00937
48599	52083	46401	62157	30185	75223	55450	39585	72003	21474	02896	93206
51040	71046	48371	12133	09991	38602	82639	99704	59214	43874	13652	61453
17371	59761	24044	04558	27686	40277	55525	94856	74298	68158	63154	55024
10226	99509	69944	25009	95236	20301	39791	43315	37310	67085	74172	45227
90943	01758	25590	11578	08990	67725	08248	39141	62266	92298	73893	41252
84641	37684	08326	24745	60827	40399	91530	26801	69585	16449	51891	32807
81579	47136	32475	33811	93580	97811	74845	19669	08245	59443	13366	41480
16198	40319	60478	14408	04620	72749	50913	69467	54069	37098	00303	00386
52759	42992	21174	89126	05001	45055	26772	31162	38035	08002	40706	19109
93974	73448	17866	80790	81070	18266	69593	05349	46019	31839	73561	28483
16546	77061	36349	85412	50954	37413	64166	77657	18889	89971	84695	37233
25013	81684	50539	09350	74826	23769	42212	12215	69337	56563	76177	70016
02819	74681	10866	52713	25350	39813	31180	93689	19375	46022	79413	93891
26998	39868	88091	58530	52226	35210	57675	49948	49879	00035	09611	77741
00393	30146	32692	84728	57231	62067	28190	24865	41114	14986	46751	66697
42807	89114	97490	93443	83658	85268	97287	08300	70623	48046	15725	91069
70002	72429	58660	89693	11575	51151	20877	67799	22396	77691	25810	66844
76935	57213	57787	78947	37795	15322	79779	64018	48728	31956	17114	82256
29164	65274	75120	50631	05970	67873	46183	69434	13665	35191	49170	02481
53796	63422	31695	32080	47715	31802	53201	47349	34451	61602	54315	61971
08920	41675	05453	94988	55166	47900	15143	85731	24881	21362	66032	51469
97898	07620	34049	33878	48882	96904	33688	17921	31956	42577	19889	44236
82445	72721	64468	73626	22497	47174	12594	99734	44765	49873	12243	43511
35894	43848	97138	00567	89017	53697	91111	49992	41983	77901	89375	70746
65542	70818	59672	49214	13188	71692	93754	99012	31905	95536	27184	51650

http

Appendix III
Generating Normal Random Numbers

In this appendix we are concerned with a method of generating random normal numbers with digital computers.

The method to be described here utilizes the Central Limit Theorem. Using this method we obtain a sequence of rational numbers $\{z_1, z_2, \ldots, z_i, \ldots\}$ in the interval $(-\infty,\infty)$ which comes from a normal population with mean μ and variance σ^2. Throughout this book we use the notation $N(\mu,\sigma^2)$ to denote a normal population with mean μ and variance σ^2. The exercises utilizing $N(\mu,\sigma^2)$ use the expression "prepare and execute a program that will take n observations from a population which is $N(\mu,\sigma^2)$." This is equivalent to performing the following algorithm *n times*:

1. Take the sum of 12 uniformly distributed random numbers $r_1, r_2, \ldots, r_{12}$.
2. Compute $X = \sigma\left(\sum_{i=1}^{12} r_i - 6\right) + \mu$.

The following is a FORTRAN subroutine of the preceding algorithm for N(0,1).

```
      FUNCTION RMS(IST)
      SUM = 0.0
      DO 2 I = 1, 12
```


http

```
    2  SUM = SUM + RCM(IST)
       RMS = SUM - 6.0
       RETURN
       END
```

In order to take observations from a population which is $N(\mu,\sigma^2)$, compute σ times RCM(IST) and then add μ. This produces a normally distributed random number with mean μ and variance σ^2. The function RMS(IST) can be called in an expression in any place where a variable is permitted. The value IST is described in Appendix II.

The table at the end of this section illustrates a set of 1000 normally distributed random numbers with mean 0 ($\mu = 0$) and variance 1 ($\sigma^2 = 1$). As with Appendix II, it is possible to read the table into the computer.

http

Normally Distributed Random Numbers with Mean 0 and Variance 1

-0,595	1,022	-0,258	-0,340	0,957	0,831	-0,201	0,632	0,078	-0,686
-1,005	1,363	0,032	0,042	-0,132	-0,669	0,611	0,827	1,872	-1,529
0,650	0,117	0,990	-0,172	0,992	0,048	-0,244	0,001	0,525	-0,435
2,155	0,912	0,899	-1,430	-1,691	0,225	-0,033	-0,373	-1,272	0,578
0,420	0,052	0,263	0,337	-0,428	0,065	-0,062	-1,413	-0,083	-0,302
1,246	-0,304	0,211	-0,891	-0,631	-0,796	1,670	0,500	-1,374	-0,911
-0,713	-1,230	0,032	-1,001	-0,062	1,368	-0,546	-1,277	-3,046	0,983
2,612	-0,081	-0,335	0,990	-0,619	-0,180	0,223	0,040	-0,584	0,208
-0,256	2,477	-1,970	-1,266	0,777	-1,430	-1,462	-0,228	0,267	-1,022
0,791	-0,073	-0,324	-0,066	1,908	-0,624	1,922	1,406	-0,811	-0,139
-1,100	1,200	-0,103	-0,716	1,226	0,695	0,581	0,529	-0,063	-0,328
-0,631	1,362	0,205	-0,756	-0,439	0,399	-0,402	1,528	0,813	0,588
1,418	1,270	-0,818	1,527	1,990	-1,325	-0,030	-0,309	-1,607	0,183
-0,908	-0,953	-0,986	-0,388	-1,100	-0,026	0,160	0,903	0,696	0,420
0,844	0,070	-0,680	0,689	-1,163	-0,412	0,351	0,532	-0,827	0,590
-2,424	1,862	0,060	-2,400	0,104	0,675	-0,706	-0,015	1,420	0,808
0,146	-0,836	0,293	0,016	0,236	-1,350	0,913	0,832	-0,643	-1,885
0,590	0,262	0,894	0,037	-1,911	2,492	-0,131	0,217	0,452	-2,105
0,231	-1,178	1,478	-0,722	0,213	0,405	-0,648	-0,369	1,419	-0,620
-1,233	0,813	0,727	0,889	-0,599	0,571	-0,634	-1,301	1,569	-0,387
0,855	-0,242	0,224	2,425	-1,379	-0,660	-0,434	-1,171	-0,661	-0,749
-1,456	0,731	0,603	-1,336	-0,558	0,476	0,730	-0,121	-0,841	-0,238
-0,050	-0,642	0,182	0,215	0,161	0,150	-1,245	0,493	-1,117	0,125
1,364	0,817	0,097	1,416	-0,203	0,407	0,879	0,182	-0,373	0,821
0,670	-1,263	-0,456	-0,353	0,558	-0,158	-0,025	-0,223	-1,650	0,887
-0,535	-0,156	0,662	1,513	-0,539	1,493	-0,317	-1,295	0,885	0,955
-0,500	0,993	-0,995	0,614	-1,307	0,351	-1,260	1,172	-1,287	0,505
-1,372	-1,344	-0,515	2,630	1,534	0,756	-1,070	-0,287	0,355	-0,170
-0,835	0,853	0,493	0,050	-0,977	-1,588	1,454	-0,987	0,210	0,249
0,012	0,465	-1,130	0,718	-0,710	0,456	-0,358	-1,617	0,518	-0,517
-0,985	-1,715	0,805	0,034	-0,325	-0,668	1,206	1,376	1,086	-0,338
-1,051	0,135	1,771	0,797	0,671	0,521	-0,246	0,295	-1,107	1,655
-1,501	0,120	-0,377	0,079	-0,900	1,376	1,385	-0,279	-0,248	0,549
1,304	1,680	1,044	0,545	-0,929	-0,733	-0,433	1,898	-0,280	-0,854
-0,215	2,509	0,775	1,425	-0,435	-1,051	0,353	2,296	2,263	-1,322
-1,516	0,893	0,276	0,506	0,127	-0,183	0,381	-0,730	-0,104	0,264
0,180	1,552	-0,366	-1,235	-1,197	0,548	-0,672	1,615	0,019	-0,792
0,058	-1,249	1,377	0,094	-0,398	0,943	0,647	0,677	-0,378	1,605
-0,395	1,526	0,181	1,445	-0,905	1,108	0,428	-0,539	-0,167	1,159
0,837	0,881	0,910	-0,980	-0,653	0,272	-0,119	1,094	1,594	-0,152
1,379	0,086	0,993	-0,936	0,986	-0,949	0,537	0,211	1,870	1,039
1,224	-1,633	0,382	0,465	0,833	-0,990	1,067	0,370	0,087	-0,105
-2,155	0,580	0,363	0,623	-0,169	1,020	-1,074	-0,601	-1,026	-0,567
-0,585	0,213	1,259	2,100	-1,280	1,038	-0,460	0,552	-1,151	2,878
-0,795	-0,309	-0,095	-0,105	0,283	0,266	-0,079	1,221	0,908	-0,549
-1,289	0,612	-0,499	-1,829	-1,263	0,818	-0,196	0,115	-0,200	1,011
0,608	-0,694	-0,633	0,598	-0,803	-0,939	0,556	-0,214	-1,336	1,087
-0,167	-0,568	-0,263	0,950	-2,133	-0,465	-0,967	0,737	-2,011	-0,770
0,114	-0,114	-0,704	0,201	-0,703	-0,767	-2,976	-2,207	-0,627	0,002
-0,094	0,478	-0,841	-0,358	-0,785	-1,664	0,753	-0,243	-2,609	0,699

http

2,334	0,064	1,641	1,130	1,865	0,216	0,458	-0,600	-0,323	0,101
-0,294	0,381	-0,775	2,034	-0,470	-0,211	1,403	-0,035	2,160	0,926
0,393	0,234	-0,545	-0,836	-0,960	0,877	-0,563	-1,517	-0,550	-0,788
-1,858	-1,724	-2,283	-0,891	0,234	-0,091	-0,611	-0,603	-0,253	-0,313
-1,550	0,230	1,026	0,063	-2,258	1,317	0,336	-0,253	-0,192	-1,298
1,593	-1,301	-1,303	-0,117	-1,155	-1,327	-1,038	0,130	0,816	-0,626
-1,277	-1,227	-0,993	0,219	-0,431	0,516	-0,128	-1,222	-0,956	0,381
1,108	-1,034	0,299	-0,221	-1,109	-0,064	0,848	0,273	0,236	-2,272
-0,019	-1,665	-1,240	0,846	0,893	0,102	-0,779	-1,651	2,361	-1,434
-1,918	-0,569	-1,038	0,589	-2,398	0,197	-0,136	-0,076	0,099	-1,594
0,968	-0,170	0,506	-0,717	-0,220	1,203	-1,913	0,917	-0,703	0,317
-0,631	1,379	0,956	-0,026	-0,604	1,257	0,933	-1,909	0,963	-0,607
0,570	0,179	0,898	0,078	-0,958	-0,186	1,095	0,548	0,546	0,699
-1,177	-0,277	-0,854	-1,065	0,408	-0,790	1,184	-1,008	1,530	-1,619
-0,157	1,274	0,788	-0,435	-0,149	0,332	1,372	0,415	0,496	-0,203
0,524	-0,077	-0,875	1,948	-0,101	1,129	1,459	0,588	0,022	-0,747
-0,060	-2,117	-0,726	1,322	1,498	-1,413	-2,168	-0,487	-1,194	-0,015
-0,695	0,273	0,993	-2,739	1,028	-0,136	-1,460	-1,185	-0,547	-0,142
-1,033	-0,708	-1,750	1,183	0,685	1,153	1,686	-1,275	0,079	0,228
1,036	-0,287	0,146	1,181	0,852	-0,651	0,017	-0,342	-0,883	-0,374
-0,683	0,964	-0,229	-2,394	0,402	-1,392	-0,552	0,344	-0,744	0,427
-1,416	-0,039	1,684	-0,504	-1,284	1,094	-0,683	0,512	2,419	-1,288
0,554	-1,086	-0,516	0,499	-0,575	1,781	-1,104	-0,955	1,722	-0,829
0,984	-0,460	0,670	1,227	1,325	1,351	0,657	-0,487	-0,869	-2,139
0,870	-0,515	0,899	-0,427	-0,703	0,452	-1,448	0,817	0,018	-2,571
-1,005	0,757	-2,365	0,015	-1,089	1,138	0,224	0,936	-2,186	-0,743
0,938	0,863	0,088	-0,650	0,153	-0,960	0,548	0,062	2,828	1,691
2,581	-1,481	-1,272	-1,973	-0,161	0,781	0,368	-0,134	-0,069	-0,005
-0,651	0,883	-2,877	-0,049	0,341	-0,422	-0,176	-1,088	-0,336	-1,504
1,987	0,275	0,429	-0,801	-1,239	-0,623	-0,921	0,797	1,553	0,613
-0,252	0,986	1,495	-0,505	-0,312	0,390	-1,096	-0,628	-2,058	0,567
-0,003	-0,370	0,395	0,138	0,210	1,198	-0,973	1,857	-0,492	0,336
-0,214	-0,081	-0,606	1,265	-0,016	-1,234	-2,340	-0,254	0,780	-0,972
0,338	1,678	0,695	1,023	1,384	0,645	0,476	-1,152	1,709	-0,016
0,441	0,326	0,382	1,074	0,103	0,531	1,005	0,572	0,642	-1,426
-1,416	-0,862	0,936	0,122	-0,888	-0,247	0,750	0,587	0,373	0,331
-1,125	0,480	1,805	-1,240	0,476	-0,610	-0,123	1,156	1,403	1,291
-1,604	-0,820	-1,499	-2,140	1,391	-0,275	-0,770	-1,441	0,181	-0,719
-0,890	1,067	-0,469	-0,381	0,800	-0,615	-0,451	0,322	0,465	-0,684
0,237	1,320	1,338	-0,269	1,881	-0,164	0,440	0,069	0,062	0,396
-0,318	1,806	0,653	-0,998	-0,575	0,056	0,687	0,393	-0,564	-1,761
-0,676	-0,559	-0,274	-1,054	0,565	-0,787	-0,804	0,329	-1,420	1,774
0,279	-1,470	-1,472	0,707	-0,182	-0,302	1,093	-1,097	1,499	1,009
-2,304	-0,162	0,737	0,755	0,511	0,367	-2,094	0,205	-2,293	1,432
-1,077	1,016	0,495	0,100	-0,265	0,838	-0,909	0,360	1,814	-0,388
0,733	1,230	-0,184	0,367	-0,555	-0,361	-0,666	-0,541	0,411	1,088
0,983	-0,860	-0,502	-0,248	0,288	-0,124	0,319	-0,177	-1,473	2,001
1,370	0,553	0,400	-0,264	0,113	0,863	-0,923	0,546	0,030	0,094
-2,397	0,215	-2,295	0,210	0,699	0,276	0,114	1,126	-0,327	-1,080
0,630	1,510	-0,583	1,998	-0,386	-0,399	0,217	-0,164	-1,769	-2,637

http

Appendix IV
Answers to Selected Examples and Exercises

http

```
C            EXERCISE 2-1
      DIMENSION IFREQ(10),CONST(11)
      DATA CONST/0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0/
      DATA IST/1346927477/,IFREQ/10*0/,ARELF/0.0/
      DATA FRESUM,FRESQ,RNDSUM,RNDSQ/4*0.0/
      DO 100 I = 1,50
      RAND=RCM(IST)
      RNDSUM=RNDSUM+RAND
      RNDSQ=RNDSQ+RAND*RAND
      J=10.0*RAND+1.0
      IFREQ(J)=IFREQ(J)+1
  100 CONTINUE
      RN=50
      WRITE(6,10)
   10 FORMAT(1H1,3X8HINTERVAL,10X9HFREQUENCY,10X
     1 18HRELATIVE FREQUENCY,10X30HACCUMULATED RELATIVE FREQUENCY)
      DO 200 K = 1,10
      FREQ=IFREQ(K)
      RELFRE=FREQ/RN
      ARELF=ARELF+RELFRE
      TEMP=CONST(K)+0.5*(CONST(K+1)-CONST(K))
      FRESUM=FRESUM+TEMP*FREQ
      FRESQ=FRESQ+TEMP*TEMP*FREQ
      WRITE(6,20)  CONST(K),CONST(K+1),IFREQ(K),RELFRE,ARELF
   20 FORMAT(1H0,3XF3.1,2H--F3.1,13XI3,19XF6.3,28XF6.3)
  200 CONTINUE
      WRITE(6,30) RNDSUM,RNDSQ
   30 FORMAT(1H0,20X26HTHE SUM OF THE NUMBERS IS F8.5,2X
     1 4HAND  F8.3, 2X25HIS THE SUM OF THE SQUARES)
      WRITE(6,40) FRESUM,FRESQ
   40 FORMAT(1H0,10X34HTHE GROUPED SUM OF THE NUMBERS IS F8.5,
     1 2X4HAND F8.5,2X33HIS THE SUM OF THE GROUPED SQUARES)
      STOP
      END
```

INTERVAL	FREQUENCY	RELATIVE FREQUENCY	ACCUMULATED RELATIVE FREQUENCY
0.0--0.1	5	0.100	0.100
0.1--0.2	2	0.040	0.140
0.2--0.3	6	0.120	0.260
0.3--0.4	7	0.140	0.400
0.4--0.5	2	0.040	0.440
0.5--0.6	5	0.100	0.540
0.6--0.7	4	0.080	0.620
0.7--0.8	9	0.180	0.800
0.8--0.9	6	0.120	0.920
0.9--1.0	4	0.080	1.000

THE SUM OF THE NUMBERS IS 26.51822 AND 18.190 IS THE SUM OF THE SQUARES

THE GROUPED SUM OF THE NUMBERS IS 26.40000 AND 17.90500 IS THE SUM OF THE GROUPED SQUARES

http

```
C          EXERCISE 2-2
      DIMENSION KTIME(2)
      REAL MEAN
      DATA KTIME/5HFIRST,6HSECOND/,IST/4536287947/,IEND/1/
      DATA N,Y,Z/100,1,0,5,0/
      SQ10=SQRT(10,0)
   50 SUMSQ=0,0
      SUM=0,0
      YSUMSQ=0,0
      YSUM=0,0
      ZSUMSQ=0,0
      ZSUM=0,0
      DO 100 I=1,N
      VALUE=RMS(IST)*SQ10+5,0
      SUM=SUM+VALUE
      SUMSQ=SUMSQ+VALUE*VALUE
      VALUE=VALUE-Y
      YSUM=YSUM+VALUE
      YSUMSQ=YSUMSQ+VALUE*VALUE
      VALUE=(VALUE+Y)*Z
      ZSUM=ZSUM+VALUE
      ZSUMSQ=ZSUMSQ+VALUE*VALUE
  100 CONTINUE
      RN=N
      MEAN=SUM/RN
      YMEAN=YSUM/RN
      ZMEAN=ZSUM/RN
      RNM1=RN-1,0
      VAR=(SUMSQ-RN*MEAN*MEAN)/RNM1
      YVAR=(YSUMSQ-RN*YMEAN*YMEAN)/RNM1
      ZVAR=(ZSUMSQ-RN*ZMEAN*ZMEAN)/RNM1
      WRITE(6,140) KTIME(IEND)
      WRITE(6,145)
      WRITE(6,150) MEAN,VAR,N
      WRITE(6,155) Y
      WRITE(6,150) YMEAN,YVAR,N
      WRITE(6,160) Z
      WRITE(6,150) ZMEAN,ZVAR,N
  140 FORMAT(1H034X8HFOR THE A6,1X24HDATA SET, THE FOLLOWING
     1 22HVALUES ARE CALCULATED )
  145 FORMAT(1H0,53X25HFOR THE UNALTERED VALUES )
  150 FORMAT(1H ,31X11HTHE MEAN IS F8,4,1X20HAND THE VARIANCE IS
     1 F8,4,1X4HFOR I3,1X13HOBSERVATIONS,)
  155 FORMAT(1H0,45X9HWITH K1= F3,1,28H, THE CALCULATED VALUES ARE )
  160 FORMAT(1H0,44X9HWITH K2= F6,4,28H, THE CALCULATED VALUES ARE )
      IF(IEND,EQ,2) STOP
      IEND = IEND + 1
      Y = 5,0
      Z = 1,0/SQ10
      GO TO 50
      END
```

```
      FOR THE FIRST  DATA SET, THE FOLLOWING VALUES ARE CALCULATED

                          FOR THE UNALTERED VALUES
   THE MEAN IS  4,9457 AND THE VARIANCE IS   9,4632 FOR 100 OBSERVATIONS,

                   WITH K1= 1.0, THE CALCULATED VALUES ARE
   THE MEAN IS  3,9457 AND THE VARIANCE IS   9,4632 FOR 100 OBSERVATIONS,

                  WITH K2= 5,0000, THE CALCULATED VALUES ARE
   THE MEAN IS 24,7287 AND THE VARIANCE IS 236,5802 FOR 100 OBSERVATIONS,

      FOR THE SECOND DATA SET, THE FOLLOWING VALUES ARE CALCULATED

                          FOR THE UNALTERED VALUES
   THE MEAN IS  4,7247 AND THE VARIANCE IS   9,3799 FOR 100 OBSERVATIONS,

                   WITH K1= 5.0, THE CALCULATED VALUES ARE
   THE MEAN IS -0,2753 AND THE VARIANCE IS   9,3799 FOR 100 OBSERVATIONS,

                  WITH K2= 0,3162, THE CALCULATED VALUES ARE
   THE MEAN IS  1,4941 AND THE VARIANCE IS   0,9380 FOR 100 OBSERVATIONS,
```


http

```
C              EXERCISE 2-3
       DIMENSION NUMB(10)
       DATA IST/453647585 1/,N,NUMB/11*0/
   50  N = N + 50
       WRITE(6,10)
   10  FORMAT(1H1)
       DO 200 I = 1,N
       NHEAD=0
       DO 100 J=1,10
       IF(RCM(IST),GT,0,5) NHEAD=NHEAD+1
  100  CONTINUE
       NUMB(NHEAD)=NUMB(NHEAD)+1
       WRITE(6,150) I,NHEAD
  150  FORMAT(1H ,40X20HON EXPERIMENT NUMBER I3,2H, I3,1X
     1 20HHEADS WERE OBSERVED.)
  200  CONTINUE
       WRITE(6,250) N
  250  FORMAT(1H1,25X4HFOR I3,1X27HEXPERIMENTS, THE FOLLOWING
     1 36HFREQUENCY DISTRIBUTION WAS OBTAINED /26X1HX5X10HNUMBER OF
     2 22HTIMES X HEADS OCCURRED,5X18HRELATIVE FREQUENCY)
       RN=N
       DO 350 I=1,10
       RELFRE=FLOAT(NUMB(I))/RN
       WRITE(6,300) I,NUMB(I),RELFRE
  300  FORMAT(1H ,24X12,19X13,26XF5,3)
       NUMB(I) = 0
  350  CONTINUE
       IF(N,LE,50) GO TO 50
       STOP
       END
```


http

```
ON EXPERIMENT NUMBER  1,    2 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER  2,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER  3,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER  4,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER  5,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER  6,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER  7,    7 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER  8,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER  9,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 10,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 11,    7 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 12,    3 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 13,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 14,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 15,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 16,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 17,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 18,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 19,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 20,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 21,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 22,    7 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 23,    3 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 24,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 25,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 26,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 27,    3 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 28,    3 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 29,    3 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 30,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 31,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 32,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 33,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 34,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 35,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 36,    7 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 37,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 38,    8 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 39,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 40,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 41,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 42,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 43,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 44,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 45,    8 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 46,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 47,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 48,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 49,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 50,    4 HEADS WERE OBSERVED.

 FOR  50 EXPERIMENTS, THE FOLLOWING FREQUENCY DISTRIBUTION WAS OBTAINED
 X      NUMBER OF TIMES X HEADS OCCURRED      RELATIVE FREQUENCY
 1                      0                          0.
 2                      1                          0.020
 3                      5                          0.100
 4                     13                          0.260
 5                     12                          0.240
 6                     13                          0.260
 7                      4                          0.080
 8                      2                          0.040
 9                      0                          0.
10                      0                          0.
```


http

```
ON EXPERIMENT NUMBER  1,    7 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER  2,    5 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER  3,    7 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER  4,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER  5,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER  6,    3 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER  7,    5 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER  8,    3 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER  9,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 10,    8 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 11,    3 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 12,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 13,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 14,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 15,    3 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 16,    7 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 17,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 18,    3 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 19,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 20,    5 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 21,    3 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 22,    3 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 23,    5 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 24,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 25,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 26,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 27,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 28,    5 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 29,    5 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 30,    3 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 31,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 32,    2 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 33,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 34,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 35,    5 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 36,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 37,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 38,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 39,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 40,    2 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 41,    8 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 42,    7 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 43,    7 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 44,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 45,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 46,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 47,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 48,    8 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 49,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 50,    6 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 51,    2 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 52,    7 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 53,    3 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 54,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 55,    7 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 56,    4 HEADS WERE OBSERVED,
ON EXPERIMENT NUMBER 57,    6 HEADS WERE OBSERVED,
```


http

```
ON EXPERIMENT NUMBER 58,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 59,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 60,    3 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 61,    7 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 62,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 63,    8 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 64,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 65,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 66,    7 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 67,    7 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 68,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 69,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 70,    3 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 71,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 72,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 73,    2 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 74,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 75,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 76,    8 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 77,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 78,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 79,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 80,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 81,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 82,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 83,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 84,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 85,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 86,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 87,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 88,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 89,    6 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 90,    2 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 91,    2 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 92,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 93,    3 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 94,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 95,    3 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 96,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 97,    1 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 98,    5 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER 99,    4 HEADS WERE OBSERVED.
ON EXPERIMENT NUMBER100,    5 HEADS WERE OBSERVED.
```

FOR 100 EXPERIMENTS, THE FOLLOWING FREQUENCY DISTRIBUTION WAS OBTAINED

X	NUMBER OF TIMES X HEADS OCCURRED	RELATIVE FREQUENCY
1	1	0.010
2	6	0.060
3	13	0.130
4	25	0.250
5	18	0.180
6	22	0.220
7	10	0.100
8	5	0.050
9	0	0.
10	0	0.

http

```
C             EXERCISE 2-4
      DIMENSION INTRVL(6),ISEG(3)
      REAL MU
      DATA MU,SIGMA,N/0.0,1.0,0/,IST/45374659731/
      SIGMA2=2.0*SIGMA
      SIGMA3=3.0*SIGMA
   50 N=N+100
      DO 75 I=1,3
      INTRVL(I)=0
      ISEG(I)=0
   75 CONTINUE
      DO 85 I=4,6
   85 INTRVL(I)=0
      WRITE(6,100)
  100 FORMAT(1H1)
      DO 400 I=1,N
      RAND=RMS(IST)
      IRAND=RAND+4.0
      IF(IRAND.LT.1.OR.IRAND.GT.6) GO TO 400
      INTRVL(IRAND)=INTRVL(IRAND)+1
      RAND=ABS(RAND)
      IF(RAND-SIGMA) 150,150,200
  150 ISEG(1)=ISEG(1)+1
  200 IF(RAND-SIGMA2) 250,250,300
  250 ISEG(2)=ISEG(2)+1
  300 ISEG(3)=ISEG(3)+1
  400 CONTINUE
      RN=N
      WRITE(6,450) N
  450 FORMAT(1H0,27X4HFOR I3,1X8HPOINTS,
     1 50HTHE FOLLOWING FREQUENCY DISTRIBUTION WAS OBTAINED /
     2 25X8HINTERVAL10X18HNUMBER IN INTERVAL10X
     3 18HRELATIVE FREQUENCY)
      DO 550 I=1,6
      RELFRE=FLOAT(INTRVL(I))/RN
      J=I-4
      K=I-3
      WRITE(6,500) J,K,INTRVL(I),RELFRE
  500 FORMAT(19X I2,1X9HSIGMA TO I2,1X5HSIGMA,11X I3,24X F5.3)
  550 CONTINUE
      IF(N.LE.100) GO TO 50
      STOP
      END
```

FOR 100 POINTS, THE FOLLOWING FREQUENCY DISTRIBUTION WAS OBTAINED

INTERVAL	NUMBER IN INTERVAL	RELATIVE FREQUENCY
-3 SIGMA TO -2 SIGMA	5	0.050
-2 SIGMA TO -1 SIGMA	11	0.110
-1 SIGMA TO 0 SIGMA	37	0.370
0 SIGMA TO 1 SIGMA	34	0.340
1 SIGMA TO 2 SIGMA	11	0.110
2 SIGMA TO 3 SIGMA	2	0.020

http

```
        FOR 200 POINTS, THE FOLLOWING FREQUENCY DISTRIBUTION WAS OBTAINED
      INTERVAL          NUMBER IN INTERVAL          RELATIVE FREQUENCY
-3 SIGMA TO -2 SIGMA              6                       0,030
-2 SIGMA TO -1 SIGMA             20                       0,100
-1 SIGMA TO  0 SIGMA             66                       0,330
 0 SIGMA TO  1 SIGMA             85                       0,425
 1 SIGMA TO  2 SIGMA             21                       0,105
 2 SIGMA TO  3 SIGMA              2                       0,010
```

Exercise 2.5

```
C               EXERCISE 2-5
      DATA IST/453769875/
   50 N=N+100
      WRITE(6,60)
   60 FORMAT(1H1)
      NHEADS=0
      DO 150 I=1,N
      IF(RCM(IST).GT.0.5) NHEADS=NHEADS+1
      IF(MOD(I,10).NE.0) GO TO 150
      RELFRE=FLOAT(NHEADS)/FLOAT(I)
      WRITE(6,100) I,NHEADS,RELFRE
  100 FORMAT(1H0,18X6HAFTER I3,1X20HTOSSES OF THE COIN, I3,
     1 1X60HHEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY
     2 3HOF F5.3)
  150 CONTINUE
      WRITE(6,200) N,RELFRE
  200 FORMAT(/////34X35HTHE FINAL RELATIVE FREQUENCY AFTER I3,
     1 1X10HTOSSES IS F5.3)
      IF(N.LE.100) GO TO 50
      STOP
      END
```

```
 AFTER  10 TOSSES OF THE COIN,   8 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.800
 AFTER  20 TOSSES OF THE COIN,  14 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.700
 AFTER  30 TOSSES OF THE COIN,  22 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.733
 AFTER  40 TOSSES OF THE COIN,  28 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.700
 AFTER  50 TOSSES OF THE COIN,  32 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.640
 AFTER  60 TOSSES OF THE COIN,  37 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.617
 AFTER  70 TOSSES OF THE COIN,  45 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.643
 AFTER  80 TOSSES OF THE COIN,  50 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.625
 AFTER  90 TOSSES OF THE COIN,  54 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.600
 AFTER 100 TOSSES OF THE COIN,  61 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.610


 THE FINAL RELATIVE FREQUENCY AFTER 100 TOSSES IS 0.610
```


http

```
AFTER  10 TOSSES OF THE COIN,   6 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.600
AFTER  20 TOSSES OF THE COIN,  12 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.600
AFTER  30 TOSSES OF THE COIN,  17 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.567
AFTER  40 TOSSES OF THE COIN,  23 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.575
AFTER  50 TOSSES OF THE COIN,  27 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.540
AFTER  60 TOSSES OF THE COIN,  32 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.533
AFTER  70 TOSSES OF THE COIN,  34 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.486
AFTER  80 TOSSES OF THE COIN,  37 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.462
AFTER  90 TOSSES OF THE COIN,  42 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.467
AFTER 100 TOSSES OF THE COIN,  46 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.460
AFTER 110 TOSSES OF THE COIN,  49 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.445
AFTER 120 TOSSES OF THE COIN,  52 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.433
AFTER 130 TOSSES OF THE COIN,  55 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.423
AFTER 140 TOSSES OF THE COIN,  60 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.429
AFTER 150 TOSSES OF THE COIN,  62 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.413
AFTER 160 TOSSES OF THE COIN,  67 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.419
AFTER 170 TOSSES OF THE COIN,  71 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.418
AFTER 180 TOSSES OF THE COIN,  76 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.422
AFTER 190 TOSSES OF THE COIN,  81 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.426
AFTER 200 TOSSES OF THE COIN,  90 HEADS WERE OBSERVED, GIVING A CUMULATIVE RELATIVE FREQUENCY OF 0.450

          THE FINAL RELATIVE FREQUENCY AFTER 200 TOSSES IS 0.450
```

Exercise 3.1

```
C                EXERCISE 3-1
      DIMENSION INUMB(6,6)
      DATA IST/5648586937/,INUMB/36*0/
      N=200
   50 DO 75 I=1,6
      DO 75 J=1,6
   75 INUMB(I,J)=0
      DO 100 I=1,N
      K=RCM(IST)*6.0+1.0
      J=RCM(IST)*6.0+1.0
      INUMB(J,K)=INUMB(J,K)+1
  100 CONTINUE
      WRITE(6,150) N
  150 FORMAT(1H1,40X28HTHE OBSERVED FREQUENCIES FOR I4,
     1 1X15HROLLS OF 2 DICE//61X9HFIRST DIE/25X1H120X1H2
     2 20X1H320X1H420X1H520X1H6)
      WRITE(6,200) (I,((INUMB(I,J)),J=1,6),I=1,6)
  200 FORMAT(11X10HSECOND DIE/6(16XI1,5XI4,5(17XI4)//))
      IF(N.GE.500) STOP
      N=500
      GO TO 50
      END
```

THE OBSERVED FREQUENCIES FOR 200 ROLLS OF 2 DICE

SECOND DIE \ FIRST DIE	1	2	3	4	5	6
1	8	4	9	8	11	3
2	4	5	8	4	7	8
3	9	8	8	6	3	3
4	3	3	2	5	7	3
5	6	6	2	5	4	4
6	6	4	8	3	8	5

THE OBSERVED FREQUENCIES FOR 500 ROLLS OF 2 DICE

SECOND DIE \ FIRST DIE	1	2	3	4	5	6
1	13	11	13	13	11	15
2	16	14	12	11	8	16
3	14	19	18	16	16	19
4	9	12	14	12	10	16
5	11	21	15	15	6	17
6	17	16	15	14	13	12

Exercise 3.4

```
C               EXERCISE 3-4
      DATA IST/5769473437/
      N=100
   50 NHEAD1=0
      NHEAD2=0
      NHD2A1=0
      DO 100 I=1,N
      J=0
      IF(RCM(IST),GT,0,4) J=1
      NHEAD1=NHEAD1+J
      K=2,0*RCM(IST)
      NHEAD2=NHEAD2+K
      NHD2A1=NHD2A1+(J+K)/2
  100 CONTINUE
      WRITE(6,150) N,NHEAD1,NHEAD2,NHD2A1
  150 FORMAT(1H011X4HFOR I3,1X26HTRIALS, COIN 1 WAS A HEAD
     1 I3,1X26HTIMES,  COIN 2 WAS A HEAD I3,
     2 1X3JHTIMES,  BOTH COINS WERE HEADS I3,1X6HTIMES,)
      IF(N,EQ,500) STOP
      N=500
      GO TO 50
      END
```

```
FOR 100 TRIALS, COIN 1 WAS A HEAD  59 TIMES,  COIN 2 WAS A HEAD  44 TIMES,  BOTH COINS WERE HEADS  25 TIMES.
FOR 500 TRIALS, COIN 1 WAS A HEAD 298 TIMES,  COIN 2 WAS A HEAD 240 TIMES,  BOTH COINS WERE HEADS 135 TIMES.
```


http

Exercise 3.5

```
C               EXERCISE 3-5
      DATA IST/7859463421/
      N=10
   50 NSTRIK=0
      NHITS=0
      NCAUT=0
      DO 300 I=1,N
      IF(RCM(IST),GT,0,8) GO TO 300
      IF(RCM(IST),GT,0,6) GO TO 200
      IF(RCM(IST),GT,0,6) GO TO 100
      NCAUT=NCAUT+1
  100 NHITS=NHITS+1
  200 NSTRIK=NSTRIK+1
  300 CONTINUE
      WRITE(6,400) N,NSTRIK,NHITS,NCAUT
  400 FORMAT(1H015X13HOF THE FIRST I3, 1X11HPLAYS, THE
     1 17HBATTER STRUCK AT I3, 1X15HBALLS, HITTING I3,
     2 1X13HOF THEM, WITH I3,1X13HBEING CAUGHT,)
      IF(N,EQ,100) STOP
      N=100
      GO TO 50
      END
```

```
OF THE FIRST  10 PLAYS, THE BATTER STRUCK AT   7 BALLS, HITTING   4 OF THEM, WITH  1 BEING CAUGHT,
OF THE FIRST 100 PLAYS, THE BATTER STRUCK AT  83 BALLS, HITTING  43 OF THEM, WITH 22 BEING CAUGHT,
```

Exercise 3.6

```
C               EXERCISE 3-6
      INTEGER R,DEN,PERMNR,COMBNR
      N=5
      R=2
   50 PERMNR=1
      I=N-R+1
      DO 100 J=I,N
      PERMNR=PERMNR*J
  100 CONTINUE
      DEN=1
      DO 200 J=1,R
      DEN=DEN*J
  200 CONTINUE
      COMBNR=PERMNR/DEN
      WRITE(6,250) N,R,PERMNR,N,R,COMBNR
  250 FORMAT(1H035X16HPERMUTATIONS OF I3,1X13HTHINGS TAKEN I3,
     1 1X12HAT A TIME IS I10,1H,/36X16HCOMBINATIONS OF I3,
     2 1X13HTHINGS TAKEN I3,1X12HAT A TIME IS I10)
      N=2*N
      IF(N,LE,20) GO TO 50
      STOP
      END
```


http

```
PERMUTATIONS OF    5 THINGS TAKEN    2 AT A TIME IS         20.
COMBINATIONS OF    5 THINGS TAKEN    2 AT A TIME IS         10

PERMUTATIONS OF   10 THINGS TAKEN    2 AT A TIME IS         90.
COMBINATIONS OF   10 THINGS TAKEN    2 AT A TIME IS         45

PERMUTATIONS OF   20 THINGS TAKEN    2 AT A TIME IS        380.
COMBINATIONS OF   20 THINGS TAKEN    2 AT A TIME IS        190
```

Exercise 3.7

```
C               EXERCISE 3-7
      INTEGER TRINGL(20,20)
      N=10
   50 DO 100 I=1,20
      DO 100 J=1,I
  100 TRINGL(I,J)=0
      DO 200 I=1,N
  200 TRINGL(I,1)=I
      DO 300 I=2,N
      DO 300 J=2,I
  300 TRINGL(I,J)=TRINGL(I-1,J)+TRINGL(I-1,J-1)
      DO 400 I=1,N
      WRITE(6,350) (TRINGL(I,J),J=1,I)
  350 FORMAT(2H 15I6,10I7,5I6)
  400 CONTINUE
      IF(N,EQ,20) STOP
      N=20
      WRITE(6,450)
  450 FORMAT(1H1)
      GO TO 50
      END
```

```
1      1
1      2      1
1      3      3      1
1      4      6      4      1
1      5     10     10      5      1
1      6     15     20     15      6      1
1      7     21     35     35     21      7      1
1      8     28     56     70     56     28      8      1
1      9     36     84    126    126     84     36      9      1
1     10     45    120    210    252    210    120     45     10      1
```

```
1    1
1    2    1
1    3    3    1
1    4    6    4    1
1    5   10   10    5    1
1    6   15   20   15    6    1
1    7   21   35   35   21    7    1
1    8   28   56   70   56   28    8    1
1    9   36   84  126  126   84   36    9    1
1   10   45  120  210  252  210  120   45   10    1
1   11   55  165  330  462  462  330  165   55   11    1
1   12   66  220  495  792  924  792  495  220   66   12    1
1   13   78  286  715 1287 1716 1716 1287  715  286   78   13    1
1   14   91  364 1001 2002 3003 3432 3003 2002 1001  364   91   14    1
1   15  105  455 1365 3003 5005 6435 6435 5005 3003 1365  455  105   15    1
1   16  120  560 1820 4368 8008 11440 12870 11440 8008 4368 1820  560  120   16    1
1   17  136  680 2380 6188 12376 19448 24310 24310 19448 12376 6188 2380  680  136   17    1
1   18  153  816 3060 8568 18564 31824 43758 48620 43758 31824 18564 8568 3060  816  153   18    1
1   19  171  969 3876 11628 27132 50388 75582 92378 92378 75582 50388 27132 11628 3876  969  171   19    1
1   20  190 1140 4845 15504 38760 77520 125970 167960 184756 167960 125970 77520 38760 15504 4845 1140  190   20    1
```

Exercise 4.1

```
C               EXERCISE 4-1
      DATA IST/7463521857/,ISUM,ISUMSQ,N/2*0,50/
      DO 200 I=1,N
      NHEADS=0
      DO 100 J=1,5
      IF(RCM(IST),GT,0,5) NHEADS=NHEADS+1
  100 CONTINUE
      ISUM=ISUM+NHEADS
      ISUMSQ=ISUMSQ+NHEADS*NHEADS
  200 CONTINUE
      RN=N
      SMEAN=FLOAT(ISUM)/RN
      SVAR=FLOAT(ISUMSQ)/RN-SMEAN*SMEAN
      WRITE(6,250) SMEAN,SVAR
  250 FORMAT(12X48HFOR THE 50 EXPERIMENTS OF TOSSING 05 COINS, THE
     1 15HSAMPLE MEAN IS F8,3, 28H AND THE SAMPLE VARIANCE IS F8,3)
      STOP
      END
```

```
FOR THE 50 EXPERIMENTS OF TOSSING 05 COINS, THE SAMPLE MEAN IS    2,520 AND THE SAMPLE VARIANCE IS    1,410
```

Exercise 4.2

```
C               EXERCISE 4-2
      DIMENSION X(2)
      INTEGER X,XX,TOTX
      DATA IST/5749731217/,X,TOTX/3*0/
      DO 100 I=1,50
      XX=0
      IF(RCM(IST),GT,0,4) XX=1
      X(XX+1)=X(XX+1)+1
      TOTX=TOTX+XX
  100 CONTINUE
      SMEAN=FLOAT(TOTX)/50,0
      SVAR=SMEAN*(1,0-SMEAN)
      FMEAN=FLOAT(X(1)*0+X(2)*1)/50,0
      FVAR=FLOAT(X(1)*0*0+X(2)*1*1)/50,0-FMEAN*FMEAN
      WRITE(6,200) SMEAN,SVAR,FMEAN,FVAR,X
  200 FORMAT(1H013X15HSAMPLE MEAN IS F8,3,
     1 19H SAMPLE VARIANCE IS F8,3,19H FREQUENCY MEAN IS F8,3,
     2 23H FREQUENCY VARIANCE IS F8,3//45X11HTHERE WERE I3,
     3 12H ZEROS, AND I3,19H ONES IN THE SAMPLE)
      STOP
      END
```

```
SAMPLE MEAN IS    0,600 SAMPLE VARIANCE IS   0,240 FREQUENCY MEAN IS    0,600 FREQUENCY VARIANCE IS    0,240
                          THERE WERE  20 ZEROS, AND  30 ONES IN THE SAMPLE
```


http

```
C               EXERCISE 4-3
      DIMENSION INTRVL(6),ISEG(3)
      DATA IST/6758493421/,INTRVL,ISEG,N/9*0,50/
      P=0.5
      SMEAN=5.0*P
      STDEV=SQRT(SMEAN*(1.0-P))
      STDV1P=SMEAN+STDEV
      STDV1M=SMEAN-STDEV
      STDV2P=STDV1P+STDEV
      STDV2M=STDV1M-STDEV
      STDV3P=STDV2P+STDEV
      STDV3M=STDV2M-STDEV
      DO 800 K=1,N
      NHEADS=0
      DO 100 J=1,5
  100 IF(RCM(IST).GT.0.5) NHEADS=NHEADS+1
      I=0
      RHEADS=NHEADS
      IF(RHEADS-STDV3M) 800,200,150
  150 IF(RHEADS-STDV2M) 200,300,250
  250 IF(RHEADS-STDV1M) 300,400,350
  350 IF(RHEADS-SMEAN)  400,500,450
  450 IF(RHEADS-STDV1P) 500,600,550
  550 IF(RHEADS-STDV2P) 600,700,650
  650 IF(STDV3P-RHEADS) 700,700,800
  700 I=1
  600 I=I+1
  500 I=I+1
  400 I=I+1
  300 I=I+1
  200 I=I+1
  750 INTRVL(I)=INTRVL(I)+1
      IF(I.LE.3) I=I+1
      L=IABS(I-4)+1
      ISEG(L)=ISEG(L)+1
  800 CONTINUE
      ISEG(2)=ISEG(2)+ISEG(1)
      ISEG(3)=ISEG(3)+ISEG(2)
      WRITE(6,900) N,INTRVL,ISEG
  900 FORMAT(30X4HFOR I3,57H TOSSES OF  5 COINS, THE FOLLOWING RESULTS
     1 WERE OBSERVED//4X26HNUMBER OF SIGMAS FROM MEAN
     2 10X2H-310X2H-210X2H-111X1H011X1H111X1H211X1H3//
     3 8X18HNUMBER IN INTERVAL 10X6(9XI3)//10X
     4 44HNUMBER IN INTERVAL -N SIGMA TO PLUS N SIGMA 33X3(I3,9X))
      STOP
      END
```

```
                    FOR  50 TOSSES OF  5 COINS, THE FOLLOWING RESULTS  WERE OBSERVED

NUMBER OF SIGMAS FROM MEAN          -3          -2          -1           0           1           2           3

   NUMBER IN INTERVAL                1           5          10          16          15           0

     NUMBER IN INTERVAL -N SIGMA TO PLUS N SIGMA                          26          46          47
```


http

Exercise 4.4

```
C               EXERCISE 4-4
      DIMENSION N(4),SMEAN(4),SVAR(4),SUM(4),SUMSQ(4)
      DATA IST/6978426517/,N/10,100,500,1000/,SUMSQ,SUM/8*0.0/
      WRITE(6,50)
   50 FORMAT(1H140X1HN14X4HMEAN10X8HVARIANCE)
      DO 300 I=1,4
      M=N(I)
      DO 100 J=1,M
      RAND=RMS(IST)
      SUM(I)=SUM(I)+RAND
      SUMSQ(I)=SUMSQ(I)+RAND*RAND
  100 CONTINUE
      RN=M
      SMEAN(I)=SUM(I)/RN
      SVAR(I)=SUMSQ(I)/RN-SMEAN(I)*SMEAN(I)
      WRITE(6,200) M,SMEAN(I),SVAR(I)
  200 FORMAT(1H038XI4,2(12XF6.3))
  300 CONTINUE
      STOP
      END
```

N	MEAN	VARIANCE
10	0,125	1,143
100	-0,157	1,020
500	0,036	1,115
1000	0,041	0,965

Exercise 4.5

```
C             EXERCISE 4-5
      DIMENSION PROB(6)
      DATA IST/7543651897/,PROB/0,33,0,41,0,20,0,05,0,01,0,00/,NPROD/0/
      DATA N/20/
      DO 50 K=2,6
   50 PROB(K)=PROB(K-1)+PROB(K)
      DO 200 I=1,N
      RAND=RCM(IST)
      DO 100 J=1,6
  100 IF(RAND,LE,PROB(J)) GO TO 150
      J=6
  150 NPROD=NPROD+J-1
  200 CONTINUE
      AVRAGE=FLOAT(NPROD)/FLOAT(N)
      WRITE(6,250) N,AVRAGE
  250 FORMAT(1H027X4HFOR I3,31H CONSECUTIVE DAYS, THE AVERAGE
     1 38HNUMBER OF PRODUCTS ACCEPTED PER DAY IS F6,3)
      STOP
      END
```

```
FOR  20 CONSECUTIVE DAYS, THE AVERAGE NUMBER OF PRODUCTS ACCEPTED PER DAY IS 0,850
```


http

```
C               EXERCISE 4-7
      INTEGER X
      DATA N,LN,M/100,0,5/
      RNMM=N-M
      RN=N
      RM=M
      P=RM/RN
  100 LN=LN+5
      RLN=LN
      CNLN=1.0
      DO 150 I=1,LN
      RI=I-1
      CNLN=CNLN*(RN-RI)/(RLN-RI)
  150 CONTINUE
      LNP1=LN+1
      DO 700 J=1,LNP1
      X=J-1
      RX=X
      CNMLNX=1.0
      LIM=LN-X
      IF(LIM.EQ.0) GO TO 180
      DO 175 K=1,LIM
      RK=K-1
      CNMLNX=CNMLNX*(RNMM-RK)/(RLN-RX-RK)
  175 CONTINUE
  180 CONTINUE
      IF(X.EQ.LNP1) CNMLNX=RNMM
      CLNX=1.0
      CMX=1.0
      IF(X.EQ.0) GO TO 300
      DO 200 K=1,X
      RK=K-1
      CMX=CMX*(RM-RK)/(RX-RK)
      CLNX=CLNX*(RLN-RK)/(RX-RK)
  200 CONTINUE
  300 HYPER=CMX*CNMLNX/CNLN
      BINOM=CLNX*(P)**X*(1.0-P)**(LN-X)
      WRITE(6,350) LN,M,X,HYPER,BINOM,P
  350 FORMAT(1H012H LITTLE N = I2,5H M = I3,
     1 5H X = I3,29H THE HYPERGEOMETRIC VALUE IS E9.3,
     2 27H AND THE BINOMIAL VALUE IS E9.3,10H WITH P = F8.5)
  700 CONTINUE
      IF(LN.LT.15) GO TO 100
      STOP
      END
```


http

LITTLE N = 5 M = 5 X = 0 THE HYPERGEOMETRIC VALUE IS 0,778E 00 AND THE BINOMIAL VALUE IS 0,774E 00 WITH P = 0.05000
LITTLE N = 5 M = 5 X = 1 THE HYPERGEOMETRIC VALUE IS 0,211E 00 AND THE BINOMIAL VALUE IS 0,204E 00 WITH P = 0.05000
LITTLE N = 5 M = 5 X = 2 THE HYPERGEOMETRIC VALUE IS 0,184E-01 AND THE BINOMIAL VALUE IS 0,214E-01 WITH P = 0.05000
LITTLE N = 5 M = 5 X = 3 THE HYPERGEOMETRIC VALUE IS 0,593E-03 AND THE BINOMIAL VALUE IS 0,113E-02 WITH P = 0.05000
LITTLE N = 5 M = 5 X = 4 THE HYPERGEOMETRIC VALUE IS 0,631E-05 AND THE BINOMIAL VALUE IS 0,297E-04 WITH P = 0.05000
LITTLE N = 5 M = 5 X = 5 THE HYPERGEOMETRIC VALUE IS 0,133E-07 AND THE BINOMIAL VALUE IS 0,312E-06 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 0 THE HYPERGEOMETRIC VALUE IS 0,584E 00 AND THE BINOMIAL VALUE IS 0,599E 00 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 1 THE HYPERGEOMETRIC VALUE IS 0,339E 00 AND THE BINOMIAL VALUE IS 0,315E 00 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 2 THE HYPERGEOMETRIC VALUE IS 0,702E-01 AND THE BINOMIAL VALUE IS 0,746E-01 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 3 THE HYPERGEOMETRIC VALUE IS 0,638E-02 AND THE BINOMIAL VALUE IS 0,105E-01 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 4 THE HYPERGEOMETRIC VALUE IS 0,251E-03 AND THE BINOMIAL VALUE IS 0,965E-03 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 5 THE HYPERGEOMETRIC VALUE IS 0,335E-05 AND THE BINOMIAL VALUE IS 0,609E-04 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 6 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,267E-05 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 7 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,804E-07 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 8 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,159E-08 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 9 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,186E-10 WITH P = 0.05000
LITTLE N = 10 M = 5 X = 10 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,977E-13 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 0 THE HYPERGEOMETRIC VALUE IS 0,436E 00 AND THE BINOMIAL VALUE IS 0,463E 00 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 1 THE HYPERGEOMETRIC VALUE IS 0,403E 00 AND THE BINOMIAL VALUE IS 0,366E 00 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 2 THE HYPERGEOMETRIC VALUE IS 0,138E 00 AND THE BINOMIAL VALUE IS 0,135E 00 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 3 THE HYPERGEOMETRIC VALUE IS 0,216E-01 AND THE BINOMIAL VALUE IS 0,307E-01 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 4 THE HYPERGEOMETRIC VALUE IS 0,154E-02 AND THE BINOMIAL VALUE IS 0,485E-02 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 5 THE HYPERGEOMETRIC VALUE IS 0,399E-04 AND THE BINOMIAL VALUE IS 0,562E-03 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 6 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,493E-04 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 7 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,334E-05 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 8 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,176E-06 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 9 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,719E-08 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 10 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,227E-09 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 11 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,543E-11 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 12 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,952E-13 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 13 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,116E-14 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 14 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,870E-17 WITH P = 0.05000
LITTLE N = 15 M = 5 X = 15 THE HYPERGEOMETRIC VALUE IS 0, AND THE BINOMIAL VALUE IS 0,305E-19 WITH P = 0.05000

http

FOR LITTLE N = 5 AND X = 0 THE BINOMIAL VALUE IS 0,774E 00 AND THE POISSON VALUE IS 0,779E 00 WITH MU = 0.250E 00
FOR LITTLE N = 5 AND X = 1 THE BINOMIAL VALUE IS 0,204E 00 AND THE POISSON VALUE IS 0,195E 00 WITH MU = 0.250E 00
FOR LITTLE N = 5 AND X = 2 THE BINOMIAL VALUE IS 0,214E-01 AND THE POISSON VALUE IS 0,243E-01 WITH MU = 0.250E 00
FOR LITTLE N = 5 AND X = 3 THE BINOMIAL VALUE IS 0,113E-02 AND THE POISSON VALUE IS 0,203E-02 WITH MU = 0.250E 00
FOR LITTLE N = 5 AND X = 4 THE BINOMIAL VALUE IS 0,297E-04 AND THE POISSON VALUE IS 0,127E-03 WITH MU = 0.250E 00
FOR LITTLE N = 5 AND X = 5 THE BINOMIAL VALUE IS 0,312E-06 AND THE POISSON VALUE IS 0,634E-05 WITH MU = 0.250E 00
FOR LITTLE N = 10 AND X = 0 THE BINOMIAL VALUE IS 0,599E 00 AND THE POISSON VALUE IS 0,607E 00 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 1 THE BINOMIAL VALUE IS 0,315E 00 AND THE POISSON VALUE IS 0,303E 00 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 2 THE BINOMIAL VALUE IS 0,746E-01 AND THE POISSON VALUE IS 0,758E-01 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 3 THE BINOMIAL VALUE IS 0,105E-01 AND THE POISSON VALUE IS 0,126E-01 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 4 THE BINOMIAL VALUE IS 0,965E-03 AND THE POISSON VALUE IS 0,158E-02 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 5 THE BINOMIAL VALUE IS 0,609E-04 AND THE POISSON VALUE IS 0,158E-03 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 6 THE BINOMIAL VALUE IS 0,267E-05 AND THE POISSON VALUE IS 0,132E-04 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 7 THE BINOMIAL VALUE IS 0,804E-07 AND THE POISSON VALUE IS 0,940E-06 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 8 THE BINOMIAL VALUE IS 0,159E-08 AND THE POISSON VALUE IS 0,588E-07 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 9 THE BINOMIAL VALUE IS 0,186E-10 AND THE POISSON VALUE IS 0,326E-08 WITH MU = 0.500E 00
FOR LITTLE N = 10 AND X = 10 THE BINOMIAL VALUE IS 0,977E-13 AND THE POISSON VALUE IS 0,163E-09 WITH MU = 0.500E 00
FOR LITTLE N = 15 AND X = 0 THE BINOMIAL VALUE IS 0,463E 00 AND THE POISSON VALUE IS 0,472E 00 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 1 THE BINOMIAL VALUE IS 0,366E 00 AND THE POISSON VALUE IS 0,354E 00 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 2 THE BINOMIAL VALUE IS 0,135E 00 AND THE POISSON VALUE IS 0,133E 00 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 3 THE BINOMIAL VALUE IS 0,307E-01 AND THE POISSON VALUE IS 0,332E-01 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 4 THE BINOMIAL VALUE IS 0,485E-02 AND THE POISSON VALUE IS 0,623E-02 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 5 THE BINOMIAL VALUE IS 0,562E-03 AND THE POISSON VALUE IS 0,934E-03 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 6 THE BINOMIAL VALUE IS 0,493E-04 AND THE POISSON VALUE IS 0,117E-03 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 7 THE BINOMIAL VALUE IS 0,334E-05 AND THE POISSON VALUE IS 0,125E-04 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 8 THE BINOMIAL VALUE IS 0,176E-06 AND THE POISSON VALUE IS 0,117E-05 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 9 THE BINOMIAL VALUE IS 0,719E-08 AND THE POISSON VALUE IS 0,977E-07 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 10 THE BINOMIAL VALUE IS 0,227E-09 AND THE POISSON VALUE IS 0,733E-08 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 11 THE BINOMIAL VALUE IS 0,543E-11 AND THE POISSON VALUE IS 0,500E-09 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 12 THE BINOMIAL VALUE IS 0,952E-13 AND THE POISSON VALUE IS 0,312E-10 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 13 HE BINOMIAL VALUE IS 0,116E-14 AND THE POISSON VALUE IS 0,180E-11 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 14 THE BINOMIAL VALUE IS 0,870E-17 AND THE POISSON VALUE IS 0,965E-13 WITH MU = 0.750E 00
FOR LITTLE N = 15 AND X = 15 THE BINOMIAL VALUE IS 0,305E-19 AND THE POISSON VALUE IS 0,483E-14 WITH MU = 0.750E 00

http

```
C              EXERCISES 4-12
      DIMENSION IFREQ(16)
      DATA IST/4839689347/,IFREQ,NB9A16,NUMB/18*0/
      DO 100 I=1,100
      J=RMS(IST)*2.0+10.0
      IF(J.LE.0) J=1
      IF(J.GT.16) J=16
      IFREQ(J)=IFREQ(J)+1
  100 CONTINUE
      DO 200 I=9,13
      NB9A16=NB9A16+IFREQ(I)
  200 CONTINUE
      WRITE(6,250) NB9A16
  250 FORMAT(1H049X34HTHE PERCENTAGE BETWEEN 9 AND 14 IS I3)
      N13A15=IFREQ(13)+IFREQ(14)
      WRITE(6,300) N13A15
  300 FORMAT(1H048X35HTHE PERCENTAGE BETWEEN 13 AND 15 IS I3)
      DO 350 I=1,2
      NUMB=NUMB+IFREQ(I)
  350 CONTINUE
      DO 450 I=3,15
      NUMB=NUMB+IFREQ(I)
      J=I+1
      RELCF=FLOAT(NUMB)/100.0
      WRITE(6,400) J,RELCF
  400 FORMAT(1H027X38HTHE RELATIVE CUMULATIVE FREQUENCY FOR
     1 23HOBSERVATIONS LESS THAN I3,3H ISF6.3)
  450 CONTINUE
      STOP
      END
```

```
                THE PERCENTAGE BETWEEN 9 AND 14 IS 66

                THE PERCENTAGE BETWEEN 13 AND 15 IS  9

THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN  4 IS 0,
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN  5 IS 0,010
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN  6 IS 0,020
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN  7 IS 0,070
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN  8 IS 0,170
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN  9 IS 0,300
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN 10 IS 0,460
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN 11 IS 0,630
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN 12 IS 0,760
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN 13 IS 0,890
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN 14 IS 0,960
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN 15 IS 0,980
THE RELATIVE CUMULATIVE FREQUENCY FOR OBSERVATIONS LESS THAN 16 IS 1,000
```


http

```
C              EXERCISE 5-1
      DIMENSION LN(2)
      REAL LCLEST,LCLACT
      DATA LN,N,RMEAN,RVAR,SSUM,SSUMSQ,IST,INTEST,INTACT/
     1 10,40,100,10,0,100,0,2*0,0,5721968731,2*0/
      SIGMA = SQRT(RVAR)
      ITIME = 1
   50 IN = LN(ITIME)
      SIN = SQRT(FLOAT(IN))
      WRITE(6,10)
   10 FORMAT(1H1,45X12HEXPERIMENT   5X4HMEAN5X8HVARIANCE)
      DO 1000 I = 1,N
      XSUM = 0.0
      XSUMSQ = 0.0
      DO 100 J = 1,IN
      TEMP = RMEAN + SIGMA*RMS(IST)
      XSUM = XSUM + TEMP
      XSUMSQ = XSUMSQ + TEMP*TEMP
  100 CONTINUE
      XMEAN = XSUM/FLOAT(IN)
      XVAR = XSUMSQ/FLOAT(IN) - XMEAN*XMEAN
      XINT = 1.96*SQRT(XVAR)/SIN
      LCLEST = XMEAN - XINT
      UCLEST = XMEAN + XINT
      XINT = 1.96*SIGMA/SIN
      LCLACT = XMEAN - XINT
      UCLACT = XMEAN + XINT
      IF(RMEAN.LE.UCLEST.AND.RMEAN.GE.LCLEST) INTEST = INTEST + 1
      IF(RMEAN.LE.UCLACT.AND.RMEAN.GE.LCLACT) INTACT = INTACT + 1
      WRITE(6,20) I,XMEAN,XVAR
   20 FORMAT(50X,I3,8X,F7.3,7X,F7.3)
      SSUM = SSUM + XSUM
      SSUMSQ = SSUMSQ + XSUMSQ
 1000 CONTINUE
      TEMP = FLOAT(N*IN)
      SMEAN = SSUM/TEMP
      SVAR = SSUMSQ/TEMP - SMEAN*SMEAN
      WRITE(6,30) N,IN,INTEST,INTACT,SMEAN,SVAR
   30 FORMAT(1H0,30X,4HFOR ,I3,18H EXPERIMENTS WITH ,I3,
     1 30H OBSERVATIONS PER EXPERIMENT, /,20X,I3,
     2 45H TIMES THE ACTUAL MEAN FELL IN THE ESTIMATED
     3 43HINTERVAL USING THE ESTIMATED VARIANCE, AND ,I3,/,50X,
     4 33H TIMES USING THE ACTUAL VARIANCE. /,30X,
     5 19HTHE SAMPLE MEAN IS ,F7.3,
     6 28H AND THE SAMPLE VARIANCE IS ,F7.3)
      ITIME = ITIME + 1
      IF(ITIME.EQ.3) STOP
      SSUM = 0.0
      SSUMSQ = 0.0
      INTEST = 0
      INTACT = 0
      GO TO 50
      END
```

EXPERIMENT	MEAN	VARIANCE
1	6,870	37,445
2	10,221	86,392
3	11,529	33,562
4	14,471	64,782
5	13,101	96,348
6	12,215	89,003
7	8,934	184,857
8	13,962	111,222
9	8,119	95,718
10	7,364	253,309
11	7,957	119,675
12	10,792	84,568
13	6,619	42,721
14	6,344	79,549
15	10,462	100,540
16	8,924	41,739
17	10,564	53,162
18	1,314	151,118
19	10,228	112,483
20	12,967	105,811
21	6,502	140,481
22	9,247	84,461
23	13,132	59,662
24	8,991	83,430
25	10,498	70,025
26	9,461	103,446
27	8,085	106,225
28	9,252	118,463
29	14,946	105,026
30	7,740	75,903
31	12,635	89,620
32	6,336	87,910
33	8,702	114,647
34	7,486	108,373
35	10,481	82,639
36	11,219	118,738
37	5,703	105,396
38	10,092	64,519
39	11,770	135,273
40	7,500	81,068
41	9,851	70,865
42	14,662	76,413
43	7,611	122,970
44	12,469	71,596
45	9,264	58,579
46	3,425	36,745
47	13,406	204,008
48	9,503	151,149
49	6,872	77,690
50	11,684	113,119
51	7,561	65,211
52	8,260	37,611
53	9,852	104,031
54	14,453	89,099
55	5,740	86,800
56	10,776	227,768
57	6,999	66,281
58	10,969	82,408
59	13,749	57,879
60	11,122	37,813
61	10,997	168,795
62	11,752	67,066
63	7,776	132,981
64	11,250	114,559
65	14,202	70,409
66	7,901	80,661
67	10,447	107,978
68	10,637	195,954
69	5,106	34,622
70	9,726	102,914
71	11,501	51,945
72	7,475	74,111
73	14,453	66,698
74	11,385	109,972
75	4,288	33,258
76	8,614	96,399
77	8,765	47,630
78	12,347	75,227
79	13,054	96,658
80	12,242	122,211
81	14,197	49,134
82	6,960	91,505
83	10,053	250,846
84	8,071	84,340
85	11,155	89,472
86	10,315	103,953
87	16,751	60,382
88	11,647	91,665
89	10,581	96,309
90	8,287	108,629

http

91	9,644	43,415
92	3,623	71,274
93	10,289	147,864
94	8,805	105,473
95	12,825	100,416
96	12,226	130,457
97	7,389	113,391
98	12,200	145,825
99	12,495	38,586
100	7,519	135,180

FOR 100 EXPERIMENTS WITH 10 OBSERVATIONS PER EXPERIMENT,
94 TIMES THE ACTUAL MEAN FELL IN THE ESTIMATED INTERVAL USING THE ESTIMATED VARIANCE, AND 96 TIMES USING THE ACTUAL VARIANCE,
THE SAMPLE MEAN IN 9,899 AND THE SAMPLE VARIANCE IN 104,115

EXPERIMENT	MEAN	VARIANCE
1	8,198	83,599
2	12,297	139,792
3	9,560	64,429
4	8,649	133,740
5	11,383	101,208
6	8,992	84,187
7	9,781	80,075
8	10,719	69,928
9	8,655	80,475
10	8,298	84,498
11	9,790	82,170
12	11,608	107,475
13	9,450	112,540
14	9,062	101,514
15	6,240	113,338
16	8,691	120,281
17	9,509	79,735
18	10,630	76,701
19	11,543	112,106
20	11,288	107,005
21	11,375	72,465
22	9,093	92,909
23	8,331	124,807
24	11,356	101,417
25	10,195	127,902
26	12,356	70,012
27	9,355	102,553
28	11,127	71,843
29	10,435	95,400
30	12,278	71,424
31	12,563	98,784
32	9,284	87,301
33	10,286	118,779
34	7,623	120,957
35	10,629	106,282
36	7,086	105,980
37	11,440	111,880
38	11,417	57,826
39	9,751	118,537
40	7,179	86,640
41	8,312	87,395
42	12,709	92,291
43	11,071	93,224
44	10,094	94,911
45	8,638	116,514
46	6,869	56,105
47	10,279	92,867
48	11,763	72,858
49	9,486	116,269
50	13,735	103,127
51	9,227	118,529
52	12,189	90,940
53	9,961	116,349
54	10,710	137,794
55	9,468	74,989
56	8,220	73,761
57	10,750	92,943
58	9,079	116,094
59	10,786	82,596
60	8,656	87,931
61	8,219	74,160
62	9,723	55,036
63	8,394	101,622
64	11,322	91,322
65	9,494	121,288
66	9,392	86,212

http

67	9.103	109.163
68	10.026	91.729
69	10.737	108.664
70	10.906	88.204
71	9.553	103.293
72	10.634	124.515
73	9.139	92.627
74	6.566	91.870
75	7.392	94.692
76	9.791	87.377
77	12.387	76.959
78	9.147	121.472
79	10.516	86.667
80	7.854	79.567
81	7.346	61.979
82	12.484	90.606
83	9.458	80.918
84	10.906	91.823
85	9.666	86.562
86	10.456	98.142
87	11.666	127.477
88	12.640	105.296
89	10.583	99.399
90	11.981	79.121
91	9.497	109.304
92	11.380	113.240
93	11.243	78.623
94	9.861	114.993
95	13.871	89.009
96	10.077	119.050
97	6.836	93.090
98	9.962	115.932
99	9.861	58.961
100	10.936	84.273

FOR 100 EXPERIMENTS WITH 40 OBSERVATIONS PER EXPERIMENT,
93 TIMES THE ACTUAL MEAN FELL IN THE ESTIMATED INTERVAL USING THE ESTIMATED VARIANCE, AND 94 TIMES USING THE ACTUAL VARIANCE.
THE SAMPLE MEAN IN 10.005 AND THE SAMPLE VARIANCE IN 98.312

```
C              EXERCISE 5-2
      DIMENSION N1(2),N2(2)
      REAL LCLEST,LCLACT
      DATA N1,N2,N,RMEAN1,RMEAN2,RVAR1,RVAR2,IST/
     1 20,40,10,30,100,5.0,10.0,1.0,1.0,7664522057/
      DIFF = RMEAN2-RMEAN1
      SIGMA1 = SQRT(RVAR1)
      SIGMA2 = SQRT(RVAR2)
      ISUB = 1
   50 INTEST = 0
      SSUM1 = 0.0
      SSUM2 = 0.0
      INTACT = 0
      SSUM = 0.0
      SSUMSQ = 0.0
      IN1 = N1(ISUB)
      IN2 = N2(ISUB)
      RIN1 = IN1
      RIN2 = IN2
      WRITE(6,10)
   10 FORMAT(1H145X12HEXPERIMENT   5X4HMEAN5X8HVARIANCE)
      DO 1000 I = 1,N
      XSUM1 = 0.0
      XSUM2 = 0.0
      XSQSM1 = 0.0
      XSQSM2 = 0.0
      DO 100 J = 1,IN1
      TEMP = RMEAN1 + SIGMA1*RMS(IST)
      XSUM1 = XSUM1 + TEMP
      XSQSM1 = XSQSM1 + TEMP*TEMP
  100 CONTINUE
      DO 200 J = 1,IN2
      TEMP = RMEAN2 + SIGMA2*RMS(IST)
      XSUM2 = XSUM2 + TEMP
      XSQSM2 = XSQSM2 + TEMP*TEMP
  200 CONTINUE
      XMEAN1 = XSUM1/RIN1
      XMEAN2 = XSUM2/RIN2
      XVAR1 = XSQSM1/RIN1 - XMEAN1*XMEAN1
      XVAR2 = XSQSM2/RIN2 - XMEAN2*XMEAN2
      XINT = 1.96*SQRT(XVAR1/RIN1 + XVAR2/RIN2)
      ESTDIF = XMEAN2 - XMEAN1
      LCLEST = ESTDIF - XINT
      UCLEST = ESTDIF + XINT
      XINT = 1.96*SQRT(RVAR1/RIN1 + RVAR2/RIN2)
      LCLACT = ESTDIF - XINT
      UCLACT = ESTDIF + XINT
      IF(DIFF.LE.UCLEST.AND.DIFF.GE.LCLEST) INTEST = INTEST + 1
      IF(DIFF.LE.UCLACT.AND.DIFF.GE.LCLACT) INTACT = INTACT + 1
      SUM = XSUM1 + XSUM2
      SUMSQ = XSQSM1 + XSQSM2
      SMEAN = SUM/(RIN1 + RIN2)
      SVAR = SUMSQ/(RIN1 + RIN2) - (RIN1*XMEAN1*XMEAN1 +
     1 RIN2*XMEAN2*XMEAN2)/(RIN1 + RIN2)
      SSUM = SSUM + SUM
      SSUMSQ = SSUMSQ + SUMSQ
      WRITE(6,20) I,SMEAN,SVAR
   20 FORMAT(50X13,8XF7.3,7XF7.3)
      SSUM1 = SSUM1 + XSUM1
      SSUM2 = SSUM2 + XSUM2
```

```
1000  CONTINUE
      TEMP = FLOAT(N*(IN1 + IN2))
      SMEAN = SSUM/TEMP
      SMEN1 = SSUM1/(RIN1*FLOAT(N))
      SMEN2 = SSUM2/(RIN2*FLOAT(N))
      SVAR = SSUMSQ/TEMP - (RIN1*SMEN1*SMEN1 + RIN2*SMEN2*SMEN2)/
     1 (RIN1 + RIN2)
      WRITE(6,30) N,IN1,IN2,INTEST,INTACT,SMEAN,SVAR
   30 FORMAT(1H040X4HFOR I3,23H EXPERIMENTS WITH N1 =
     1 I3,10H AND N2 = I3/20XI3,4H AND I4,
     2 50H ARE THE NUMBER OF TIMES THE DIFFERENCE OF MEANS
     3 30HFELL IN THE COMPUTED INTERVAL,/38XF7.3,
     4 17H IS THE MEAN, AND F7.3, 17H IS THE VARIANCE,)
      ISUB = ISUB + 1
      IF(ISUB.EQ.3) STOP
      GO TO 50
      END
```

EXPERIMENT	MEAN	VARIANCE
1	6.709	0.949
2	6.473	1.510
3	6.702	1.550
4	6.731	0.687
5	6.665	1.383
6	6.664	0.646
7	6.630	0.393
8	6.707	0.833
9	6.699	1.012
10	6.850	1.200
11	6.813	1.411
12	6.684	1.037
13	6.666	0.945
14	6.246	0.723
15	6.736	1.315
16	6.894	0.879
17	6.628	0.879
18	6.462	0.835
19	6.813	1.076
20	6.530	0.574
21	6.542	1.565
22	6.762	0.440
23	6.198	0.771
24	6.494	0.821
25	6.331	0.588
26	6.756	1.136
27	6.866	1.456
28	6.772	1.057
29	6.583	1.170
30	6.875	1.175
31	6.637	0.614
32	6.647	0.801
33	6.551	0.900
34	6.668	1.026
35	6.567	1.255
36	6.604	0.661
37	6.726	0.646
38	7.039	1.613
39	6.424	0.383
40	7.078	0.926
41	7.011	0.806
42	6.500	0.927
43	6.637	0.663
44	7.189	1.050
45	6.514	0.548
46	6.356	0.606

http

47	6,639	1,052
48	6,765	1,092
49	6,644	0,669
50	6,030	0,715
51	6,647	0,438
52	6,558	0,815
53	6,332	0,764
54	6,804	0,849
55	6,764	1,026
56	6,606	0,785
57	6,502	0,731
58	6,703	0,571
59	6,790	0,725
60	6,931	0,711
61	6,998	0,949
62	6,775	1,402
63	6,742	1,136
64	6,592	0,724
65	6,609	1,366
66	6,575	0,923
67	6,440	0,884
68	6,656	0,886
69	6,854	0,994
70	6,634	0,672
71	7,034	0,564
72	6,851	0,644
73	6,842	0,951
74	6,645	0,727
75	6,444	0,877
76	6,566	1,012
77	6,630	0,821
78	6,543	1,000
79	6,580	0,855
80	6,570	1,065
81	6,743	0,643
82	6,396	0,892
83	6,339	1,133
84	6,919	0,925
85	6,642	0,938
86	6,627	1,280
87	6,817	1,183
88	6,454	0,747
89	6,724	0,675
90	6,630	0,706
91	6,617	0,921
92	6,777	0,683
93	6,748	0,749
94	6,507	0,665
95	6,512	0,824
96	7,062	0,635
97	6,837	0,801
98	6,846	0,659
99	6,719	1,358
100	6,679	0,609

FOR 100 EXPERIMENTS WITH N1 = 20 AND N2 = 10
94 AND 96 ARE THE NUMBER OF TIMES THE DIFFERENCE OF MEANS FELL IN THE COMPUTED INTERVAL,
6,665 IS THE MEAN, AND 0,963 IS THE VARIANCE,

http

EXPERIMENT	MEAN	VARIANCE
1	7,270	0,811
2	7,092	1,233
3	7,137	0,982
4	7,237	1,228
5	7,186	0,926
6	7,094	1,027
7	7,220	1,281
8	7,197	1,210
9	7,212	0,777
10	7,097	1,072
11	7,002	0,913
12	7,412	0,758
13	7,191	1,046
14	7,035	0,780
15	7,209	0,719
16	7,306	0,947
17	7,248	1,082
18	6,888	0,807
19	7,036	0,940
20	7,026	1,485
21	7,153	1,003
22	7,057	0,902
23	7,113	0,906
24	7,132	1,016
25	7,032	1,032
26	7,144	0,923
27	7,098	1,184
28	7,146	0,869
29	7,303	0,792
30	7,367	0,999
31	7,170	0,955
32	6,993	1,034
33	7,231	0,970
34	6,960	1,194
35	7,200	0,769
36	7,165	1,140
37	7,410	1,074
38	7,074	0,994
39	7,143	1,012
40	7,040	1,185
41	7,191	1,325
42	7,287	1,288
43	7,223	1,364
44	7,157	1,158
45	7,141	1,110
46	7,251	0,936
47	7,127	1,122
48	6,928	1,024
49	7,098	0,852
50	7,082	0 906
51	7,017	0,922
52	7,190	0,914
53	7,027	1,212
54	7,088	1,115
55	7,195	0.888
56	6,957	1,006
57	7,202	1,353
58	6,992	0,891
59	7,063	0,709
60	7,176	1,100
61	7,226	0,814
62	7,061	0,962
63	6,972	1,216
64	7,200	1,181
65	7,087	0.911
66	7,085	1,040
67	7,271	1,091
68	7,111	1,408
69	7,235	1,518
70	7,448	0,966
71	7,412	0,864
72	7,211	0,890
73	6,945	1,055
74	7,166	1,121
75	7,258	0,955
76	7,212	0,886
77	7,172	1,082

http

78	7,330	0,934
79	6,962	0,669
80	7,213	0,855
81	6,999	1,228
82	7,190	1,190
83	7,206	0,758
84	7,158	0,990
85	7,198	0,899
86	7,270	1,182
87	6,948	0,889
88	7,283	1,073
89	6,955	1,047
90	7,254	0,976
91	7,067	1,096
92	6,969	0,877
93	7,168	1,076
94	7,122	0,860
95	7,343	1,076
96	7,277	1,212
97	7,242	0,946
98	7,265	0,885
99	7,319	0,942
100	7,052	1,274

FOR 100 EXPERIMENTS WITH N1 = 40 AND N2 = 30
98 AND 98 ARE THE NUMBER OF TIMES THE DIFFERENCE OF MEANS FELL IN THE COMPUTED INTERVAL.
7,153 IS THE MEAN, AND 1,050 IS THE VARIANCE.

Exercise 5.3

```
C              EXERCISE 5-3
      DATA N,RMEAN,RVAR,XSUM,XSUMSQ,IST/16,50.0,100.0,2*0.0,
     1 9752341567/
       STD=SQRT(RVAR)
       DO 100 I=1,N
       TEMP=STD*RMS(IST)+RMEAN
       XSUM=XSUM+TEMP
       XSUMSQ=XSUMSQ+TEMP*TEMP
  100  CONTINUE
       AVE=XSUM/FLOAT(N)
       SVAR=XSUMSQ/FLOAT(N)-AVE*AVE
       DEN=SQRT(RVAR/FLOAT(N))
       TEST40=(AVE-40.0)/DEN
       TEST49=(AVE-49.0)/DEN
       TEST50=(AVE-50.0)/DEN
       TEST51=(AVE-51.0)/DEN
       TEST60=(AVE-60.0)/DEN
       WRITE(6,200) AVE,SVAR,TEST40,TEST49,TEST50,TEST51,TEST60
  200  FORMAT(1H0,47X28HTHE COMPUTED SAMPLE MEAN IS F6.2/
     1 1H0,45X32HTHE COMPUTED SAMPLE VARIANCE IS F6.2/
     2 5(/55X16HA TEST VALUE IS F6.2),//43X
     3 44HTHE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1.96)
       STOP
       END
```

```
      THE COMPUTED SAMPLE MEAN IS  52.57

   THE COMPUTED SAMPLE VARIANCE IS  92.76

            A TEST VALUE IS   5.03
            A TEST VALUE IS   1.43
            A TEST VALUE IS   1.03
            A TEST VALUE IS   0.63
            A TEST VALUE IS  -2.97

   THE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1.96
```


http

```
C              EXERCISE 5-4
      DATA N,RMEAN,RVAR,XSUM,XSUMSQ,IST/500,50.0,100.0,2*0.0,
     1 5648965431/
      STD=SQRT(RVAR)
      DO 100 I=1,N
      TEMP=STD*RMS(IST)+RMEAN
      XSUM=XSUM+TEMP
      XSUMSQ=XSUMSQ+TEMP*TEMP
  100 CONTINUE
      AVE=XSUM/FLOAT(N)
      SVAR=XSUMSQ/FLOAT(N)-AVE*AVE
      DEN=SQRT(RVAR/FLOAT(N))
      TEST40=(AVE-40.0)/DEN
      TEST49=(AVE-49.0)/DEN
      TEST51=(AVE-51.0)/DEN
      TEST50=(AVE-50.0)/DEN
      TEST60=(AVE-60.0)/DEN
      WRITE(6,200) AVE,SVAR,TEST40,TEST49,TEST50,TEST51,TEST60
  200 FORMAT(1H0,47X28HTHE COMPUTED SAMPLE MEAN IS F6.2/
     1 1H0,45X32HTHE COMPUTED SAMPLE VARIANCE IS F6.2/
     2 5(/55X16HA TEST VALUE IS F6.2),//43X
     3 44HTHE SIGNIFICANCE LEVEL FOR A 1% TEST IS 2.58)
      STOP
      END
```

```
          THE COMPUTED SAMPLE MEAN IS  50.34

       THE COMPUTED SAMPLE VARIANCE IS 101.78

                A TEST VALUE IS  23.12
                A TEST VALUE IS   3.00
                A TEST VALUE IS   0.76
                A TEST VALUE IS  -1.48
                A TEST VALUE IS -21.60

      THE SIGNIFICANCE LEVEL FOR A 1% TEST IS 2.58
```


http

Exercise 5.5

```
C                EXERCISE 5-5
      DATA N,RMEAN,RVAR,XSUM,XSUMSQ,IST/25,50.0,100.0,2*0.0,
     1 6194265431/
      STD=SQRT(RVAR)
      DO 100 I=1,N
      TEMP=STD*RMS(IST)+RMEAN
      XSUM=XSUM+TEMP
      XSUMSQ=XSUMSQ+TEMP*TEMP
  100 CONTINUE
      AVE=XSUM/FLOAT(N)
      SVAR=XSUMSQ/FLOAT(N)-AVE*AVE
      DEN=SQRT(RVAR/FLOAT(N))
      TEST49=(AVE-49.0)/DEN
      TEST50=(AVE-50.0)/DEN
      WRITE(6,200) AVE,SVAR,TEST49,TEST50
  200 FORMAT(1H0,47X28HTHE COMPUTED SAMPLE MEAN IS F6.2/
     1 1H0,45X32HTHE COMPUTED SAMPLE VARIANCE IS F6.2/
     2 2(/55X16HA TEST VALUE IS F6.2),//
     3 15X44HTHE SIGNIFICANCE LEVEL FOR A 1% TEST IS 2.58 5X
     4 44HTHE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1.96)
      STOP
      END
```

```
                THE COMPUTED SAMPLE MEAN IS  47.42

              THE COMPUTED SAMPLE VARIANCE IS 109.93

                      A TEST VALUE IS  -0.79
                      A TEST VALUE IS  -1.29

THE SIGNIFICANCE LEVEL FOR A 1% TEST IS 2.58     THE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1.96
```


http

```
C              EXERCISE 5-6
      DATA N1,N2,IST,RMEAN1,RMEAN2,RVAR1,RVAR2,SUM1,SUM2,
     1 SUMSQ1,SUMSQ2/20,25,6749204751,2*50.0,100.0,25.0,
     2 4*0.0/
       SIGMA1 = SQRT(RVAR1)
       SIGMA2 = SQRT(RVAR2)
       DO 100 I = 1,N1
       TEMP = RMEAN1 + SIGMA1*RMS(IST)
       SUM1 = SUM1 + TEMP
       SUMSQ1 = SUMSQ1 + TEMP*TEMP
  100  CONTINUE
       DO 200 I = 1,N2
       TEMP = RMEAN2 + SIGMA2*RMS(IST)
       SUM2 = SUM2 + TEMP
       SUMSQ2 = SUMSQ2 + TEMP*TEMP
  200  CONTINUE
       RN1 = N1
       RN2 = N2
       SMEAN1 = SUM1/RN1
       SMEAN2 = SUM2/RN2
       SVAR1 = SUMSQ1/RN1 - SMEAN1*SMEAN1
       SVAR2 = SUMSQ2/RN2 - SMEAN2*SMEAN2
       DEN = SQRT(SVAR1/RN1 + SVAR2/RN2)
       DIFF = SMEAN2 - SMEAN1
       DM10 = (DIFF + 10.0)/DEN
       DM1 = (DIFF + 1.0)/DEN
       D0 = DIFF/DEN
       D1 = (DIFF - 1.0)/DEN
       D10 = (DIFF - 10.0)/DEN
       WRITE(6,10) SMEAN1,SMEAN2,SVAR1,SVAR2,DM10,DM1,D0,D1,D10
   10  FORMAT(1H040X17HSAMPLE MEANS ARE  2(5XF7.3)/
     1 38X21HSAMPLE VARIANCES ARE  2(5XF7.3)/
     2 5(55X13HA Z-VALUE IS  F7.3/),45X
     3 45HTHE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1.96 )
       STOP
       END
```

```
   SAMPLE MEANS ARE        49.841       50.666
SAMPLE VARIANCES ARE       98.777       23.459
                    A Z-VALUE IS   4.465
                    A Z-VALUE IS   0.753
                    A Z-VALUE IS   0.340
                    A Z-VALUE IS  -0.072
                    A Z-VALUE IS  -3.785
         THE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1.96
```


http

```
C              EXERCISE 5-7
      DATA N1,N2,IST,RMEAN1,RMEAN2,RVAR1,RVAR2,SUM1,SUM2,
     1 SUMSQ1,SUMSQ2/200,500,67492047S1,2*50.0,100.0,25.0,
     2 4*0.0/
       SIGMA1 = SQRT(RVAR1)
       SIGMA2 = SQRT(RVAR2)
       DO 100 I = 1,N1
       TEMP = RMEAN1 + SIGMA1*RMS(IST)
       SUM1 = SUM1 + TEMP
       SUMSQ1 = SUMSQ1 + TEMP*TEMP
  100  CONTINUE
       DO 200 I = 1,N2
       TEMP = RMEAN2 + SIGMA2*RMS(IST)
       SUM2 = SUM2 + TEMP
       SUMSQ2 = SUMSQ2 + TEMP*TEMP
  200  CONTINUE
       RN1 = N1
       RN2 = N2
       SMEAN1 = SUM1/RN1
       SMEAN2 = SUM2/RN2
       SVAR1 = SUMSQ1/RN1 - SMEAN1*SMEAN1
       SVAR2 = SUMSQ2/RN2 - SMEAN2*SMEAN2
       DEN = SQRT(SVAR1/RN1 + SVAR2/RN2)
       DIFF = SMEAN2 - SMEAN1
       DM10 = (DIFF + 10.0)/DEN
       DM1 = (DIFF + 1.0)/DEN
       D0 = DIFF/DEN
       D1 = (DIFF - 1.0)/DEN
       D10 = (DIFF - 10.0)/DEN
       WRITE(6,10) SMEAN1,SMEAN2,SVAR1,SVAR2,DM10,DM1,D0,D1,D10
   10  FORMAT(1H040X17HSAMPLE MEANS ARE  2(5XF7.3)/
     1 38X21HSAMPLE VARIANCES ARE  2(5XF7.3)/
     2 5(55X13HA Z-VALUE IS  F7.3/),45X
     3 45HTHE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1.96 )
       STOP
       END
```

```
   SAMPLE MEANS ARE        49,347       49,644
SAMPLE VARIANCES ARE       86,015       24,068
                  A Z-VALUE IS  14,890
                  A Z-VALUE IS   1,876
                  A Z-VALUE IS   0,430
                  A Z-VALUE IS  -1,017
                  A Z-VALUE IS -14,031
        THE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1,96
```


http

```
C               EXERCISE 5-8
      DATA N1,N2,IST,RMEAN1,RMEAN2,RVAR1,RVAR2,SUM1,SUM2,
     1 SUMSQ1,SUMSQ2/100,25,6749204751,2*50.0,100.0,25.0,
     2 4*0.0/
        SIGMA1 = SQRT(RVAR1)
        SIGMA2 = SQRT(RVAR2)
        DO 100 I = 1,N1
        TEMP = RMEAN1 + SIGMA1*RMS(IST)
        SUM1 = SUM1 + TEMP
        SUMSQ1 = SUMSQ1 + TEMP*TEMP
  100   CONTINUE
        DO 200 I = 1,N2
        TEMP = RMEAN2 + SIGMA2*RMS(IST)
        SUM2 = SUM2 + TEMP
        SUMSQ2 = SUMSQ2 + TEMP*TEMP
  200   CONTINUE
        RN1 = N1
        RN2 = N2
        SMEAN1 = SUM1/RN1
        SMEAN2 = SUM2/RN2
        SVAR1 = SUMSQ1/RN1 - SMEAN1*SMEAN1
        SVAR2 = SUMSQ2/RN2 - SMEAN2*SMEAN2
        DEN = SQRT(SVAR1/RN1 + SVAR2/RN2)
        DIFF = SMEAN2 - SMEAN1
        DM1 = (DIFF + 1.0)/DEN
        D0 = DIFF/DEN
        D1 = (DIFF - 1.0)/DEN
        WRITE(6,10) SMEAN1,SMEAN2,SVAR1,SVAR2,DM1,D0,D1
   10   FORMAT(1H040X17HSAMPLE MEANS ARE  2(5XF7.3)/
     1 38X21HSAMPLE VARIANCES ARE  2(5XF7.3)/
     2 3(55X13HA Z-VALUE IS  F7.3/),45X
     3 45HTHE SIGNIFICANCE LEVEL FOR A 1% TEST IS 2.58 /45X
     4 45HTHE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1.96  )
        STOP
        END
```

```
   SAMPLE MEANS ARE        50.658       48.688
SAMPLE VARIANCES ARE       91.190        15.170
                  A Z-VALUE IS  -0.787
                  A Z-VALUE IS  -1.599
                  A Z-VALUE IS  -2.410
        THE SIGNIFICANCE LEVEL FOR A 1% TEST IS 2.58
        THE SIGNIFICANCE LEVEL FOR A 5% TEST IS 1.96
```


http

```
C              EXERCISE 6-1
      DIMENSION TALPHA(2),LN(2)
      REAL LCLEST
      DATA LN,TALPHA,N,RMEAN,RVAR,SSUM,SSUMSQ,INTEST,IST/10,40,2,262,
     1 1.96,100,10.0,100.0,2*0.0,0,7434764727/
      SIGMA = SQRT(RVAR)
      ITIME = 1
   50 IN = LN(ITIME)
      SIN = SQRT(FLOAT(IN))
      WRITE(6,10)
   10 FORMAT(1H1,45X,12HEXPERIMENT #,5X,4HMEAN,5X,8HVARIANCE)
      DO 1000 I = 1,N
      XSUM = 0.0
      XSUMSQ = 0.0
      DO 100 J = 1,IN
      TEMP = RMEAN + SIGMA*RMS(IST)
      XSUM = XSUM + TEMP
      XSUMSQ = XSUMSQ + TEMP*TEMP
  100 CONTINUE
      XMEAN = XSUM/FLOAT(IN)
      XVAR = XSUMSQ/FLOAT(IN) - XMEAN*XMEAN
      XINT = TALPHA(ITIME)*SQRT(XVAR)/SIN
      LCLEST = XMEAN - XINT
      UCLEST = XMEAN + XINT
      IF(RMEAN.LE.UCLEST.AND.RMEAN.GE.LCLEST) INTEST = INTEST + 1
      WRITE(6,20) I,XMEAN,XVAR
   20 FORMAT(50X,I3,8X,F7.3,7X,F7.3)
      SSUM = SSUM + XSUM
      SSUMSQ = SSUMSQ + XSUMSQ
 1000 CONTINUE
      TEMP = FLOAT(N*IN)
      SMEAN = SSUM/TEMP
      SVAR = SSUMSQ/TEMP - SMEAN*SMEAN
      WRITE(6,30) N,IN,INTEST,SMEAN,SVAR
   30 FORMAT(1H0,30X,4HFOR ,I3,18H EXPERIMENTS WITH ,I3,30H OBSERVATIONS PE
     1R EXPERIMENT, /20X,I3,83H TIMES THE ACTUAL MEAN FELL IN THE ESTIMAT
     2ED INTERVAL USING THE ESTIMATED VARIANCE./30X,19HTHE SAMPLE MEAN IS
     3 F7.3,28H AND THE SAMPLE VARIANCE IS F7.3)
      ITIME = ITIME + 1
      IF(ITIME.EQ.3) STOP
      SSUM = 0.0
      SSUMSQ = 0.0
      INTEST = 0
      GO TO 50
      END
```


http

EXPERIMENT #	MEAN	VARIANCE
1	9,418	69,763
2	12,507	72,832
3	0,065	61,979
4	8,180	43,646
5	10,704	65,655
6	15,013	131,334
7	12,332	40,650
8	9,456	57,967
9	13,916	87,959
10	14,477	73,926
11	9,998	77,748
12	5,765	72,105
13	6,791	86,674
14	10,848	93,471
15	7,065	44,636
16	9,754	41,586
17	14,507	241,862
18	8,609	60,421
19	8,991	126,894
20	11,134	113,807
21	12,168	69,353
22	4,303	112,607
23	8,489	68,305
24	9,197	62,809
25	8,115	140,140
26	10,626	143,103
27	14,155	62,798
28	8,809	85,185
29	11,578	200,271
30	6,385	78,053
31	10,413	107,999
32	9,545	98,979
33	14,084	33,099
34	12,126	62,956
35	11,702	73,537
36	17,949	118,885
37	8,476	172,397
38	6,686	66,479
39	11,279	34,769
40	10,419	100,543
41	4,838	166,225
42	11,619	109,924
43	14,687	100,911
44	9,984	46,147
45	11,941	40,126
46	9,677	35,369
47	10,887	20,640
48	11,381	160,919
49	9,249	77,770
50	9,225	88,988
51	17,266	54,953
52	9,323	18,126
53	8,311	98,536
54	13,418	53,537
55	4,868	113,736
56	15,362	51,301
57	11,270	104,437
58	12,222	32,998
59	14,126	142,323
60	15,180	73,470
61	14,842	242,629
62	12,415	85,815
63	7,310	32,890
64	9,500	111,420
65	7,762	27,567
66	8,048	188,625
67	7,179	153,301
68	6,148	103,003
69	5,337	94,354
70	7,072	114,769
71	12,630	56,460
72	10,173	99,642
73	5,881	90,339
74	12,727	94,699
75	12,604	32,765
76	12,653	84,342
77	5,546	50,110

http

78	11,993	100,748
79	9,190	58,838
80	13,466	88,126
81	13,389	128,999
82	11,544	72,770
83	8,092	75,716
84	14,278	97,540
85	10,542	132,511
86	11,952	76,390
87	7,819	116,692
88	10,848	48,779
89	10,453	49,509
90	9,183	66,908
91	14,178	68,686
92	5,676	49,154
93	10,245	81,327
94	12,838	37,880
95	8,774	143,310
96	10,132	145,707
97	12,191	113,915
98	14,428	124,817
99	4,337	109,940
100	15,021	89,147

FOR 100 EXPERIMENTS WITH 10 OBSERVATIONS PER EXPERIMENT,
96 TIMES THE ACTUAL MEAN FELL IN THE ESTIMATED INTERVAL USING THE ESTIMATED VARIANCE.
THE SAMPLE MEAN IS 10,393 AND THE SAMPLE VARIANCE IS 98,772

EXPERIMENT #	MEAN	VARIANCE
1	9,756	113,487
2	9,580	139,971
3	10,357	91,339
4	9,535	69,229
5	8,978	78,453
6	11,767	91,311
7	11,382	135,731
8	8,507	122,727
9	11,869	90,730
10	9,918	90,854
11	10,512	133,900
12	10,930	94,000
13	11,240	142,223
14	8,374	67,402
15	10,032	61,260
16	7,984	123,044
17	10,097	85,240
18	10,695	153,591
19	12,290	70,078
20	12,907	55,981
21	8,500	106,524
22	9,284	69,924
23	12,799	110,955
24	11,223	94,124
25	10,619	89,403
26	11,610	103,134
27	9,505	98,377
28	11,855	103,656
29	9,954	127,896
30	4,513	125,212
31	13,302	111,822
32	8,897	118,557
33	8,323	103,995
34	9,302	129,460
35	11,135	70,046
36	9,742	103,307
37	7,533	94,929
38	6,656	72,333
39	9,558	97,313
40	9,535	128,290
41	8,473	114,652
42	11,275	130,489
43	8,173	81,860
44	10,115	119,833
45	9,837	108,164
46	10,709	98,039
47	10,915	53,166

http

48	9.545	75.415
49	9.143	112.375
50	11.572	123.417
51	12.362	98.700
52	12.097	131.138
53	9.357	53.110
54	12.011	73.787
55	8.097	91.063
56	10.419	146.146
57	5.049	105.713
58	12.895	77.452
59	9.895	82.024
60	10.071	102.820
61	11.418	135.124
62	10.696	125.557
63	10.644	83.486
64	10.954	121.630
65	8.975	84.275
66	12.218	76.034
67	12.679	114.250
68	11.059	123.223
69	10.338	89.769
70	9.326	104.036
71	11.229	84.874
72	10.909	86.864
73	8.560	112.180
74	11.974	119.216
75	8.093	74.198
76	10.895	77.502
77	10.987	111.877
78	8.177	114.811
79	8.137	88.659
80	8.250	129.066
81	7.405	72.579
82	9.432	100.874
83	9.668	92.199
84	8.321	92.318
85	10.914	116.362
86	8.536	86.780
87	10.512	92.756
88	7.949	59.428
89	9.097	102.191
90	12.868	81.443
91	8.603	97.802
92	11.686	89.381
93	11.602	61.226
94	10.681	106.340
95	10.948	75.445
96	9.357	117.690
97	10.019	116.128
98	8.129	50.643
99	9.944	77.721
100	9.940	78.639

FOR 100 EXPERIMENTS WITH 40 OBSERVATIONS PER EXPERIMENT,
93 TIMES THE ACTUAL MEAN FELL IN THE ESTIMATED INTERVAL USING THE ESTIMATED VARIANCE.
THE SAMPLE MEAN IS 10.037 AND THE SAMPLE VARIANCE IS 101.375

http

```
C              EXERCISE 6-2
      DIMENSION N1(2),N2(2),TALPHA(2)
      REAL LCLEST
      DATA N1,N2,N,RMEAN1,RMEAN2,RVAR1,RVAR2,IST/
     1 20,40,10,30,100,5.0,10.0,1.0,1.0,7664522057/
      DATA TALPHA/2.045,1.96/
      DIFF = RMEAN2-RMEAN1
      SIGMA1 = SQRT(RVAR1)
      SIGMA2 = SQRT(RVAR2)
      ISUB = 1
   50 INTEST = 0
      SSUM1 = 0.0
      SSUM2 = 0.0
      SSUM = 0.0
      SSUMSQ = 0.0
      IN1 = N1(ISUB)
      IN2 = N2(ISUB)
      RIN1 = IN1
      RIN2 = IN2
      WRITE(6,10)
   10 FORMAT(1H145X12HEXPERIMENT   5X4HMEAN5X8HVARIANCE)
      DO 1000 I = 1,N
      XSUM1 = 0.0
      XSUM2 = 0.0
      XSQSM1 = 0.0
      XSQSM2 = 0.0
      DO 100 J = 1,IN1
      TEMP = RMEAN1 + SIGMA1*RMS(IST)
      XSUM1 = XSUM1 + TEMP
      XSQSM1 = XSQSM1 + TEMP*TEMP
  100 CONTINUE
      DO 200 J = 1,IN2
      TEMP = RMEAN2 + SIGMA2*RMS(IST)
      XSUM2 = XSUM2 + TEMP
      XSQSM2 = XSQSM2 + TEMP*TEMP
  200 CONTINUE
      XMEAN1 = XSUM1/RIN1
      XMEAN2 = XSUM2/RIN2
      XVAR1 = XSQSM1/RIN1 - XMEAN1*XMEAN1
      XVAR2 = XSQSM2/RIN2 - XMEAN2*XMEAN2
      XVAR = (RIN1*XVAR1 + RIN2*XVAR2)/(RIN1 + RIN2 - 2.0)
      XINT = TALPHA(ISUB)*SQRT(XVAR*(1.0/RIN1 + 1.0/RIN2))
      ESTDIF = XMEAN2 - XMEAN1
      LCLEST = ESTDIF - XINT
      UCLEST = ESTDIF + XINT
      IF(DIFF.LE.UCLEST.AND.DIFF.GE.LCLEST) INTEST = INTEST + 1
      SUM = XSUM1 + XSUM2
      SUMSQ = XSQSM1 + XSQSM2
      SMEAN = SUM/(RIN1 + RIN2)
      SVAR = SUMSQ/(RIN1 + RIN2) - (RIN1*XMEAN1*XMEAN1 +
```


http

```
     1 RIN2*XMEAN2*XMEAN2)/(RIN1 + RIN2)
       SSUM = SSUM + SUM
       SSUMSQ = SSUMSQ + SUMSQ
       WRITE(6,20) I,SMEAN,SVAR
  20   FORMAT(50XI3,8XF7.3,7XF7.3)
       SSUM1 = SSUM1 + XSUM1
       SSUM2 = SSUM2 + XSUM2
1000   CONTINUE
       TEMP = FLOAT(N*(IN1 + IN2))
       SMEAN = SSUM/TEMP
       SMEN1 = SSUM1/(RIN1*FLOAT(N))
       SMEN2 = SSUM2/(RIN2*FLOAT(N))
       SVAR = SSUMSQ/TEMP - (RIN1*SMEN1*SMEN1 + RIN2*SMEN2*SMEN2)/
     1 (RIN1 + RIN2)
       WRITE(6,30) N,IN1,IN2,INTEST,SMEAN,SVAR
  30   FORMAT(1H040X4HFOR I3,23H EXPERIMENTS WITH N1 =
     1 I3,10H AND N2 = I3/20XI3,
     2 50H ARE THE NUMBER OF TIMES THE DIFFERENCE OF MEANS
     3 30HFELL IN THE COMPUTED INTERVAL,/38XF7.3,
     4 17H IS THE MEAN, AND F7.3, 17H IS THE VARIANCE,)
       ISUB = ISUB + 1
       IF(ISUB.EQ.3) STOP
       GO TO 50
       END
```


http

EXPERIMENT	MEAN	VARIANCE
1	6,709	0,949
2	6,473	1,513
3	6,702	1,550
4	6,731	0,687
5	6,665	1,383
6	6,664	0,646
7	6,630	0,393
8	6,707	0,833
9	6,699	1,012
10	6,850	1,200
11	6,813	1,411
12	6,684	1,037
13	6,666	0,945
14	6,246	0,723
15	6,736	1,315
16	6,894	0,879
17	6,628	0,879
18	6,462	0,835
19	6,813	1,076
20	6,530	0,574
21	6,542	1,565
22	6,762	0,440
23	6,198	0,771
24	6,494	0,821
25	6,331	0,588
26	6,756	1,136
27	6,866	1,456
28	6,772	1,057
29	6,583	1,170
30	6,875	1,175
31	6,637	0,614
32	6,647	0,801
33	6,551	0,900
34	6,668	1,026
35	6,567	1,255
36	6,604	0,661
37	6,726	0,646
38	7,039	1,613
39	6,424	0,383
40	7,078	0,926
41	7,011	0,806
42	6,500	0,927
43	6,637	0,663
44	7,189	1,050
45	6,514	0,548
46	6,356	0,606
47	6,639	1,052
48	6,765	1,092
49	6,644	0,669
50	6,030	0,715
51	6,647	0,438
52	6,558	0,815
53	6,332	0,784
54	6,804	0,849
55	6,764	1,026
56	6,606	0,785
57	6,502	0,731
58	6,703	0,571
59	6,790	0,725
60	6,931	0,711
61	6,998	0,949
62	6,775	1,402
63	6,742	1,136
64	6,592	0,724
65	6,609	1,366
66	6,575	0,923
67	6,440	0,884
68	6,656	0,886
69	6,854	0,994
70	6,634	0,672
71	7,034	0,564
72	6,851	0,644
73	6,842	0,951
74	6,645	0,727
75	6,444	0,877
76	6,566	1,012
77	6,630	0,821

http

78	6,543	1,000
79	6,580	0,855
80	6,570	1,065
81	6,743	0,643
82	6,396	0,892
83	6,339	1,133
84	6,919	0,925
85	6,642	0,938
86	6,627	1,280
87	6,817	1,163
88	6,454	0,747
89	6,724	0,675
90	6,630	0,706
91	6,617	0,921
92	6,777	0,683
93	6,748	0,749
94	6,507	0,665
95	6,512	0,824
96	7,062	0,635
97	6,837	0,801
98	6,846	0,659
99	6,719	1,358
100	6,679	0,609

FOR 100 EXPERIMENTS WITH N1 = 20 AND N2 = 10
96 ARE THE NUMBER OF TIMES THE DIFFERENCE OF MEANS FELL IN THE COMPUTED INTERVAL,
6,665 IS THE MEAN, AND 0,963 IS THE VARIANCE,

EXPERIMENT	MEAN	VARIANCE
1	7,270	0,811
2	7,092	1,233
3	7,137	0,982
4	7,237	1,228
5	7,186	0,926
6	7,094	1,027
7	7,220	1,281
8	7,197	1,210
9	7,212	0,777
10	7,097	1,072
11	7,002	0,913
12	7,412	0,758
13	7,191	1,046
14	7,035	0,780
15	7,209	0,719
16	7,306	0,947
17	7,248	1,082
18	6,888	0,807
19	7,036	0,940
20	7,026	1,485
21	7,153	1,003
22	7,057	0,902
23	7,113	0,906
24	7,132	1,016
25	7,032	1,032
26	7,144	0,923
27	7,098	1,184
28	7,146	0,869
29	7,303	0,792
30	7,367	0,999
31	7,170	0,955
32	6,993	1,034
33	7,231	0,970
34	6,960	1,194
35	7,200	0,769
36	7,165	1,140
37	7,410	1,074
38	7,074	0,994
39	7,143	1,012
40	7,040	1,185
41	7,191	1,325
42	7,287	1,288

http

43	7,223	1,364
44	7,157	1,158
45	7,141	1,110
46	7,251	0,936
47	7,127	1,122
48	6,928	1,024
49	7,098	0,852
50	7,082	0,906
51	7,017	0,922
52	7,190	0,914
53	7,027	1,212
54	7,088	1,115
55	7,195	0,888
56	6,957	1,006
57	7,202	1,353
58	6,992	0,891
59	7,063	0,709
60	7,176	1,100
61	7,226	0,814
62	7,061	0,962
63	6,972	1,216
64	7,200	1,181
65	7,087	0,911
66	7,085	1,040
67	7,271	1,091
68	7,111	1,408
69	7,235	1,518
70	7,448	0,966
71	7,412	0,864
72	7,211	0,890
73	6,945	1,055
74	7,166	1,121
75	7,258	0,955
76	7,212	0,886
77	7,172	1,082
78	7,330	0,934
79	6,962	0,669
80	7,213	0,855
81	6,999	1,228
82	7,190	1,190
83	7,206	0,758
84	7,158	0,990
85	7,198	0,899
86	7,270	1,182
87	6,948	0,889
88	7,283	1,073
89	6,955	1,047
90	7,254	0,976
91	7,067	1,096
92	6,969	0,877
93	7,168	1,076
94	7,122	0,860
95	7,343	1,076
96	7,277	1,212
97	7,242	0,946
98	7,265	0,885
99	7,319	0,942
100	7,052	1,274

FOR 100 EXPERIMENTS WITH N1 = 40 AND N2 = 30
98 ARE THE NUMBER OF TIMES THE DIFFERENCE OF MEANS FELL IN THE COMPUTED INTERVAL,
7,153 IS THE MEAN, AND 1,050 IS THE VARIANCE

http

```
C               EXERCISE 6-3
      DATA N,RMEAN,RVAR,XSUM,XSUMSQ,IST/10,50.0,100.0,2*0.0,
     1 9482165431/
        STD=SQRT(RVAR)
        DO 100 I=1,N
        TEMP=STD*RMS(IST)+RMEAN
        XSUM=XSUM+TEMP
        XSUMSQ=XSUMSQ+TEMP*TEMP
  100   CONTINUE
        AVE=XSUM/FLOAT(N)
        SVAR=XSUMSQ/FLOAT(N)-AVE*AVE
        DEN=SQRT(RVAR/FLOAT(N))
        TEST40=(AVE-40.0)/DEN
        TEST49=(AVE-49.0)/DEN
        TEST50=(AVE-50.0)/DEN
        TEST51=(AVE-51.0)/DEN
        TEST60=(AVE-60.0)/DEN
        WRITE(6,200) AVE,SVAR,TEST40,TEST49,TEST50,TEST51,TEST60
  200   FORMAT(1H0,47X28HTHE COMPUTED SAMPLE MEAN IS F6.2/
     1 1H0,45X32HTHE COMPUTED SAMPLE VARIANCE IS F6.2/
     2 5(/55X16HA TEST VALUE IS F6.2))
        STOP
        END
```

```
THE COMPUTED SAMPLE MEAN IS  50.09

THE COMPUTED SAMPLE VARIANCE IS 100.49

A TEST VALUE IS   3.19
A TEST VALUE IS   0.34
A TEST VALUE IS   0.03
A TEST VALUE IS  -0.29
A TEST VALUE IS  -3.13
```


http

```
C              EXERCISE 6-4
      DIMENSION DIFFMU(3),SUM1(3),SUMSQ1(3),SUM2(3),SUMSQ2(3),TEST(3),
     1 SMEAN1(3),SMEAN2(3),SVAR1(3),SVAR2(3)
      EQUIVALENCE (SUM1,SMEAN1),(SUM2,SMEAN2),(SUMSQ1,SVAR1),(SUMSQ2,
     1 SVAR2)
      DATA N1,N2,DIFFMU,SUM1,SUM2,SUMSQ1,SUMSQ2,IST,RMEAN,RVAR/6,7,0,0,
     1 1,0,20,0,12*0,0,5254893049,50,0,100,0/,D/0,0/
      SIGMA = SQRT(RVAR)
      DO 100 I = 1,N1
      TEMP = RMEAN + SIGMA*RMS(IST)
      DO 100 J = 1,3
      ATEMP = TEMP - DIFFMU(J)
      SUM1(J) = SUM1(J) + ATEMP
      SUMSQ1(J) = SUMSQ1(J) + ATEMP*ATEMP
  100 CONTINUE
      DO 200 I = 1,N2
      TEMP = RMEAN + SIGMA*RMS(IST)
      DO 200 J = 1,3
      ATEMP = TEMP + DIFFMU(J)
      SUM2(J) = SUM2(J) + ATEMP
      SUMSQ2(J) = SUMSQ2(J) + ATEMP*ATEMP
  200 CONTINUE
      RN1 = N1
      RN2 = N2
      DO 300 I = 1,3
      SMEAN1(I) = SUM1(I)/RN1
      SMEAN2(I) = SUM2(I)/RN2
      SVAR1(I) = SUMSQ1(I)/RN1 - SMEAN1(I)*SMEAN1(I)
      SVAR2(I) = SUMSQ2(I)/RN2 - SMEAN2(I)*SMEAN2(I)
      TEST(I) = (SMEAN1(I) - SMEAN2(I) - D)/SQRT(((RN1-1.0)*SVAR1(I)
     1 + (RN2-1.0)*SVAR2(I))*(1.0/RN1 + 1.0/RN2)/(RN1 + RN2 - 2.0))
  300 CONTINUE
      WRITE(6,10)
   10 FORMAT(1H022X8HDATA SET10X5HMEAN15X5HMEAN210X9HVARIANCE15X9HVARIA
     1NCE210X10HTEST VALUE)
      WRITE(6,20)(I,SMEAN1(I),SMEAN2(I),SVAR1(I),SVAR2(I),TEST(I)),
     1 I = 1,3)
   20 FORMAT(1H025X12,13XF5.2,5XF5.2,11XF6.2, 9XF6.2,13XF6.3)
      WRITE(6,30)
   30 FORMAT(1H033X68HTHE CRITICAL T-VALUE FOR A 5% TEST OF 11 DEGREES
     1OF FREEDOM IS 2.201)
      STOP
      END
```

DATA SET	MEAN1	MEAN2	VARIANCE1	VARIANCE2	TEST VALUE
1	48.80	48.88	38.82	35.14	-0.024
2	47.80	49.88	38.82	35.14	-0.617
3	28.80	68.88	38.82	35.14	11.874

THE CRITICAL T-VALUE FOR A 5% TEST OF 11 DEGREES OF FREEDOM IS 2.201

http

```
C              EXERCISE 6-5
      REAL LCLEST
      DATA IN,N,RMEAN,RVAR,SSUM,SSUMSQ,INTEST,IST/10,
     1 100,10,0,100,0,2*0,0,0,7434764727/
      DATA CHISQ,CHISUM,CSUMSQ/3*0,0/
      DATA CHIL,CHIU/2,70,19,02/
      SIGMA = SQRT(RVAR)
      RIN = IN
      SIN = SQRT(RIN)
      WRITE(6,10)
   10 FORMAT(1H138X12HEXPERIMENT  5X4HMEAN7X8HVARIANCE8X10HCHI-SQUARE )
      DO 1000 I = 1,N
      XSUM = 0.0
      XSUMSQ = 0.0
      DO 100 J = 1,IN
      TEMP = RMEAN + SIGMA*RMS(IST)
      XSUM = XSUM + TEMP
      XSUMSQ = XSUMSQ + TEMP*TEMP
  100 CONTINUE
      XMEAN = XSUM/FLOAT(IN)
      XVAR = XSUMSQ/FLOAT(IN) - XMEAN*XMEAN
      LCLEST = (RIN-1.0)*XVAR/CHIU
      UCLEST = (RIN-1.0)*XVAR/CHIL
      CHISQ = (RIN-1.0)*XVAR/RVAR
      IF(RVAR.LE.UCLEST.AND.RVAR.GE.LCLEST) INTEST = INTEST + 1
      WRITE(6,20) I,XMEAN,XVAR,CHISQ
   20 FORMAT(42X13,8XF7.3,7XF7.3,7XF7.3)
      SSUM = SSUM + XSUM
      SSUMSQ = SSUMSQ + XSUMSQ
      CHISUM = CHISUM + CHISQ
      CSUMSQ = CSUMSQ + CHISQ*CHISQ
 1000 CONTINUE
      TEMP = FLOAT(N*IN)
      SMEAN = SSUM/TEMP
      SVAR = SSUMSQ/TEMP - SMEAN*SMEAN
      WRITE(6,30) N,IN,INTEST,SMEAN,SVAR
   30 FORMAT(1H030X4HFOR I3,18H EXPERIMENTS WITH I3,30H OBSERVATIONS PE
     1R EXPERIMENT, /20X13,87H TIMES THE ACTUAL VARIANCE FELL IN THE EST
     2IMATED INTERVAL USING THE ESTIMATED VARIANCE,/30X19HTHE SAMPLE MEA
     3N IS F7.3,28H AND THE SAMPLE VARIANCE IS F7.3)
      CHIAVE = CHISUM/FLOAT(N)
      CHIVAR = CSUMSQ/FLOAT(N) - CHIAVE*CHIAVE
      WRITE(6,40) CHIAVE,CHIVAR
   40 FORMAT(34X23HAVERAGE OF CHI-SQUARES F7.3,5X
     1 24HVARIANCE OF CHI-SQUARES F7.3)
      STOP
      END
```

EXPERIMENT	MEAN	VARIANCE	CHI-SQUARE
1	9.418	69.763	6.279
2	12.507	72.832	6.555
3	0.065	61.979	5.578
4	8.180	43.646	3.928
5	10.704	65.655	5.909
6	15.013	131.334	11.820
7	12.332	40.650	3.658
8	9.456	57.967	5.217
9	13.916	87.959	7.916
10	14.477	73.926	6.653
11	9.998	77.748	6.997
12	5.765	72.105	6.489
13	6.791	86.674	7.801
14	10.848	93.471	8.412
15	7.065	44.636	4.017
16	9.754	41.586	3.743
17	14.507	241.862	21.768
18	8.609	60.421	5.438
19	8.991	126.894	11.420
20	11.134	113.807	10.243
21	12.168	69.353	6.242
22	4.303	112.607	10.135
23	8.489	68.305	6.147
24	9.197	62.809	5.653
25	8.115	140.140	12.613
26	10.626	143.103	12.879
27	14.155	62.798	5.652
28	8.809	85.185	7.667
29	11.578	200.271	18.024

http

30	6,385	78,053	7,025
31	10,413	107,999	9,720
32	9,545	98,979	8,908
33	14,084	33,099	2,979
34	12,126	62,956	5,666
35	11,702	73,537	6,618
36	17,949	118,885	10,700
37	8,476	172,397	15,516
38	6,686	66,479	5,983
39	11,279	34,769	3,129
40	10,419	100,543	9,049
41	4,838	166,225	14,960
42	11,619	109,924	9,893
43	14,687	100,911	9,082
44	9,984	46,147	4,153
45	11,941	40,126	3,611
46	9,677	35,369	3,183
47	10,887	20,640	1,858
48	11,381	160,819	14,474
49	9,249	77,770	6,999
50	9,225	88,988	8,009
51	17,266	54,953	4,946
52	9,323	18,126	1,631
53	8,311	98,536	8,868
54	13,418	53,537	4,818
55	4,868	113,736	10,236
56	15,362	51,301	4,617
57	11,270	104,437	9,399
58	12,222	32,998	2,970
59	14,126	142,323	12,809
60	15,180	73,470	6,612
61	14,842	242,629	21,837
62	12,415	85,815	7,723
63	7,310	32,890	2,960
64	9,500	111,420	10,028
65	7,762	27,567	2,481
66	8,048	188,625	16,976
67	7,179	153,301	13,797
68	6,148	103,003	9,270
69	5,337	94,354	8,492
70	7,072	114,769	10,329
71	12,630	56,460	5,081
72	10,173	99,642	8,968
73	5,881	90,339	8,130
74	12,727	94,699	8,523
75	12,604	32,765	2,949
76	12,653	84,342	7,591
77	5,546	50,110	4,510
78	11,993	100,748	9,067
79	9,190	58,838	5,295
80	13,466	88,126	7,931
81	13,389	128,999	11,610
82	11,544	72,770	6,549
83	8,092	75,716	6,814
84	14,278	97,540	8,779
85	10,542	132,511	11,926
86	11,952	76,390	6,875
87	7,819	116,692	10,502
88	10,848	48,779	4,390
89	10,453	49,509	4,456
90	9,183	66,908	6,022
91	14,178	68,686	6,182
92	5,676	49,154	4,424
93	10,245	81,327	7,319
94	12,838	37,880	3,409
95	8,774	143,310	12,898
96	10,132	145,707	13,114
97	12,191	113,915	10,252
98	14,428	124,817	11,234
99	4,337	109,940	9,895
100	15,021	89,147	8,023

FOR 100 EXPERIMENTS WITH 10 OBSERVATIONS PER EXPERIMENT,
95 TIMES THE ACTUAL VARIANCE FELL IN THE ESTIMATED INTERVAL USING THE ESTIMATED VARIANCE,
THE SAMPLE MEAN IS 10,393 AND THE SAMPLE VARIANCE IS 98,772
AVERAGE OF CHI-SQUARES 7,999 VARIANCE OF CHI-SQUARES 15,491

```
C              EXERCISE 6-6
       DATA N,IST/8,8753219547/,RMEAN,RVAR/50.0,100.0/,SUM,SUMSQ/2*0.0/
       SIGMA = SQRT(RVAR)
       DO 100 I = 1,N
       TEMP = RMEAN + SIGMA*RMS(IST)
       SUM = SUM + TEMP
       SUMSQ = SUMSQ + TEMP*TEMP
  100  CONTINUE
       RN = N
       SMEAN = SUM/RN
       SVAR = SUMSQ/RN - SMEAN*SMEAN
       CHI10 = (RN - 1.0)*SVAR/10.0
       CHI99 = (RN - 1.0)*SVAR/99.0
       CHI100 = (RN - 1.0) *SVAR/100.0
       CHI101 = (RN - 1.0)*SVAR/101.0
       CHI1000 = (RN - 1.0)*SVAR/1000.0
       WRITE(6,10) SVAR,CHI10,CHI99,CHI100,CHI101,CHI1000
   10  FORMAT(1H053X23HTHE SAMPLE VARIANCE IS F7.3// 5(43X37HA CHI-SQUAR
     1E STATISTIC CALCULATED IS F7.3//))
       STOP
       END
```

```
THE SAMPLE VARIANCE IS  45.960

A CHI-SQUARE STATISTIC CALCULATED IS  32.172

A CHI-SQUARE STATISTIC CALCULATED IS   3.250

A CHI-SQUARE STATISTIC CALCULATED IS   3.217

A CHI-SQUARE STATISTIC CALCULATED IS   3.185

A CHI-SQUARE STATISTIC CALCULATED IS   0.322
```


http

```
C              EXERCISE 6-7
      DIMENSION VARMUL(3),SUM2(3),SUMSQ2(3),
     1 SMEAN2(3),SVAR2(3),F(3)
      EQUIVALENCE (SUM2,SMEAN2),(SUMSQ2,SVAR2)
      DATA N1,N2,VARMUL,SUM1,SUM2,SUMSQ1,SUMSQ2,IST,RMEAN1,RMEAN2,RVAR
     1 8,9,1,0,1,1,5,0,8*0.0,7639812547,40.0,50.0,100.0/
      DATA C/0.0/
      SIGMA = SQRT(RVAR)
      DO 100 I = 1,N1
      TEMP = RMEAN1 + SIGMA*RMS(IST)
      SUM1 = SUM1 + TEMP
      SUMSQ1 = SUMSQ1 + TEMP*TEMP
  100 CONTINUE
      DO 200 I = 1,N2
      TEMP = RMEAN2 + SIGMA*RMS(IST)
      DO 200 J = 1,3
      ATEMP = TEMP*VARMUL(J)
      SUM2(J) = SUM2(J) + ATEMP
      SUMSQ2(J) = SUMSQ2(J) + ATEMP*ATEMP
  200 CONTINUE
      RN1 = N1
      RN2 = N2
      SMEAN1 = SUM1/RN1
      SVAR1 = SUMSQ1/RN1 - SMEAN1*SMEAN1
      DO 300 I = 1,3
      SMEAN2(I) = SUM2(I)/RN2
      SVAR2(I) = SUMSQ2(I)/RN2 - SMEAN2(I)*SMEAN2(I)
      F(I) = (SVAR1/(RN1 - 1.0))/(SVAR2(I)/(RN2 - 1.0))
  300 CONTINUE
      WRITE(6,10)
   10 FORMAT(1H022X8HDATA SET10X9HVARIANCE15X9HVARIANCE
     1 10X11HF-STATISTIC)
      WRITE(6,20)(I,SVAR1,SVAR2(I),F(I)), I = 1,3)
   20 FORMAT(1H025XI2,12XF8.3,7XF8.3,11XF6.2)
      WRITE(6,30)
   30 FORMAT(1H018X75HTHE CRITICAL F-VALUE FOR A 5( TEST WITH 7 AND 8
     1DEGREES OF FREEDOM IS 3.50 )
      STOP
      END
```

DATA SET	VARIANCE1	VARIANCE	F-STATISTIC
1	59.133	101.177	0.67
2	59.133	122.424	0.55
3	59.133	2529.422	0.03

THE CRITICAL F-VALUE FOR A 5(TEST WITH 7 AND 8 DEGREES OF FREEDOM IS 3.50

http

```
C               EXERCISE 7-1
      DIMENSION N(2),ALPHA(2),BETA(2),NDIF(2)
      DATA N,ALPHA,BETA,NDIF/50,200,5.0,4.0,1.0,1.5,1,4/
      DATA IST/8675431667/
      DO 1000 K = 1,2
      DO 900 I = 1,2
      NN = N(I)
      RNN = NN
      SUMX = 0.0
      SUMY = 0.0
      SUMERR = 0.0
      SUMXY = 0.0
      SUMXSQ = 0.0
      SUMYSQ = 0.0
      WRITE(6,700)
      DO 300 J = 1,NN
      X = 0.0
      IF(J.GT.NN/2) X = 1.0
      TEMP = RMS(IST)
      Y = ALPHA(K) + BETA(K)*X + TEMP
      SUMY = SUMY + Y
      SUMX = SUMX + X
      SUMERR = SUMERR + TEMP
      IF(J/NDIF(I)*NDIF(I).EQ.J) WRITE(6,800) X,Y
      SUMXY = SUMXY + X*Y
      SUMXSQ = SUMXSQ + X*X
      SUMYSQ = SUMYSQ + Y*Y
  300 CONTINUE
      XBAR = SUMX/RNN
      YBAR = SUMY/RNN
      ERRBAR = SUMERR/RNN
      BETAC = (RNN*SUMXY - SUMX*SUMY)/(RNN*SUMXSQ - SUMX*SUMX)
      ALPHAC = YBAR - BETAC*XBAR
      SSE = SUMYSQ - RNN*YBAR*YBAR - BETAC/RNN*
     1 (RNN*SUMXY - SUMX*SUMY)
      SVAR = SSE/(RNN - 2.0)
      WRITE(6,600) YBAR,ALPHAC,BETAC,ERRBAR,SVAR
  600 FORMAT(1H0 5(10XF10.3)//)
  700 FORMAT(1H1 52X27H50 DATA POINTS AS X-Y PAIRS)
  800 FORMAT(1H  55X1H(F4.2,1H,F5.2,1H))
  900 CONTINUE
 1000 CONTINUE
      STOP
      END
```

50 DATA POINTS AS X-Y PAIRS

(0, , 4,36)
(0. , 4,90)
(0, , 3,73)
(0, , 4,91)
(0, , 5,53)
(0, , 5,04)
(0, , 6,42)
(0, , 4,57)
(0, , 5,18)
(0, , 4,05)
(0, , 4,47)
(0, , 4,45)
(0, , 2,03)
(0, , 5,76)
(0, , 5,21)
(0, , 3,90)
(0, , 5,14)
(0, , 5,98)
(0, , 6,53)
(0, , 5,06)
(0, , 4,07)
(0, , 6,30)
(0, , 4,08)
(0, , 3,67)
(0, , 4,64)
(1,00, 5,06)
(1,00, 5,36)
(1,00, 4,07)
(1,00, 4,64)
(1,00, 7,31)
(1,00, 5,79)
(1,00, 6,58)
(1,00, 7,42)
(1,00, 5,84)
(1,00, 5,34)
(1,00, 5,89)
(1,00, 6,28)
(1,00, 6,21)
(1,00, 6,46)
(1,00, 5,08)
(1,00, 3,25)
(1,00, 8,03)
(1,00, 7,42)
(1,00, 6,29)
(1,00, 5,95)
(1,00, 6,52)
(1,00, 5,06)
(1,00, 5,94)
(1,00, 6,90)
(1,00, 5,31)

5,360	4,799	1,122	-0.140	1,103

http

50 DATA POINTS AS X-Y PAIRS

(0. , 5.38)
(0. , 5.69)
(0. , 5.66)
(0. , 5.50)
(0. , 4.77)
(0. , 3.44)
(0. , 5.00)
(0. , 5.12)
(0. , 5.18)
(0. , 4.20)
(0. , 5.26)
(0. , 6.57)
(0. , 3.17)
(0. , 3.88)
(0. , 6.05)
(0. , 5.24)
(0. , 4.99)
(0. , 5.75)
(0. , 5.67)
(0. , 4.15)
(0. , 4.53)
(0. , 5.30)
(0. , 5.51)
(0. , 4.96)
(0. , 5.85)
(1.00, 5.59)
(1.00, 6.27)
(1.00, 5.42)
(1.00, 6.97)
(1.00, 6.51)
(1.00, 6.34)
(1.00, 6.68)
(1.00, 7.02)
(1.00, 5.41)
(1.00, 5.38)
(1.00, 4.98)
(1.00, 5.94)
(1.00, 7.23)
(1.00, 7.32)
(1.00, 6.43)
(1.00, 5.50)
(1.00, 6.01)
(1.00, 7.11)
(1.00, 5.94)
(1.00, 4.36)
(1.00, 5.78)
(1.00, 5.99)
(1.00, 6.00)
(1.00, 5.38)
(1.00, 5.40)

5,456	4,970	0,974	-0,044	0,905

http

50 DATA POINTS AS X-Y PAIRS

(0, , 4,18)
(0. , 4,73)
(0, , 2,79)
(0, , 4,94)
(0, , 4,60)
(0, , 4,45)
(0, , 4,32)
(0, , 3,69)
(0, , 3,22)
(0, , 5,05)
(0, , 4,87)
(0, , 3,80)
(0, , 4,57)
(0, , 5,23)
(0, , 2,26)
(0, , 4,21)
(0. , 4,07)
(0, , 4,82)
(0, , 4,11)
(0, , 3,50)
(0, , 4,48)
(0, , 2,27)
(0, , 4,10)
(0, , 3,86)
(0. , 2,16)
(1,00, 4,04)
(1,00, 7,98)
(1,00, 5,68)
(1,00, 6,66)
(1,00, 6,12)
(1,00, 5,02)
(1,00, 6,41)
(1,00, 5,80)
(1,00, 4,06)
(1,00, 5,08)
(1,00, 5,66)
(1,00, 3,80)
(1,00, 4,78)
(1,00, 5,43)
(1,00, 5,25)
(1,00, 4,95)
(1,00, 4,68)
(1,00, 7,03)
(1,00, 4,90)
(1,00, 5,21)
(1,00, 4,51)
(1,00, 4,02)
(1,00, 6,64)
(1,00, 4,65)
(1,00, 7,13)

4,716	4,011	1.409	-0,034	0,982

http

50 DATA POINTS AS X-Y PAIRS

(0, , 4,74)
(0, , 3,08)
(0, , 3,79)
(0, , 3,57)
(0, , 2,72)
(0, , 3,19)
(0, , 3,45)
(0, , 4,78)
(0, , 3,57)
(0, , 3,64)
(0, , 3,27)
(0, , 4,89)
(0, , 2,97)
(0, , 3,89)
(0, , 4,75)
(0, , 3,20)
(0, , 6,16)
(0, , 4,25)
(0, , 3,37)
(0, , 3,39)
(0, , 2,78)
(0, , 4,11)
(0, , 2,89)
(0, , 5,18)
(0, , 3,02)
(1,00, 4,90)
(1,00, 4,79)
(1,00, 4,90)
(1,00, 4,15)
(1,00, 6,70)
(1,00, 5,93)
(1,00, 4,56)
(1,00, 5,53)
(1,00, 5,86)
(1,00, 5,31)
(1,00, 4,78)
(1,00, 5,15)
(1,00, 6,10)
(1,00, 4,26)
(1,00, 6,01)
(1,00, 5,05)
(1,00, 7,00)
(1,00, 5,32)
(1,00, 4,37)
(1,00, 6,68)
(1,00, 4,03)
(1,00, 6,00)
(1,00, 6,22)
(1,00, 5,33)
(1,00, 5,35)

4,844	4,109	1,469	0,094	1,049

http

```
C               EXERCISE 7-2
      DIMENSION BETA(3)
      DATA NN,ALPHA,BETA,BETEST/10,5.0,0.0,0.0,1.0,3.0,0.0/
      DATA IST/8675431667/
      RNN = NN
      DO 1000 K = 1,3
      SUMX = 0.0
      SUMY = 0.0
      SUMXY = 0.0
      SUMXSQ = 0.0
      SUMYSQ = 0.0
      DO 300 J = 1,NN
      X = 1.0
      IF(J.GT.NN/2) X = 2.0
      Y = ALPHA + BETA(K)*X + RMS(IST)
      SUMY = SUMY + Y
      SUMX = SUMX + X
      SUMXY = SUMXY + X*Y
      SUMXSQ = SUMXSQ + X*X
      SUMYSQ = SUMYSQ + Y*Y
  300 CONTINUE
      XBAR = SUMX/RNN
      YBAR = SUMY/RNN
      BETAC = (RNN*SUMXY - SUMX*SUMY)/(RNN*SUMXSQ - SUMX*SUMX)
      ALPHAC = YBAR - BETAC*XBAR
      SSE = SUMYSQ - RNN*YBAR*YBAR - BETAC/RNN*
     1 (RNN*SUMXY - SUMX*SUMY)
      SVAR = SSE/(RNN - 2.0)
      T = (BETAC - BETEST)*SQRT((SUMXSQ - RNN*XBAR*XBAR)/SVAR)
      WRITE(6,600) T
  600 FORMAT(1H055X13HA T-VALUE IS F7.3)
 1000 CONTINUE
      STOP
      END
```

```
A T-VALUE IS   0.741

A T-VALUE IS   2.490

A T-VALUE IS   3.806
```


http

```
C               EXERCISE 7-3
      DIMENSION NS(2),TALPHA(2)
      DATA ALPHA,BETA0,NS,N/5.0,1.0,10,40,100/
      DATA IST/8675431667/
      DATA TALPHA/2.262,1.96/
      DO 900 I = 1,2
      NN = NS(I)
      INTEST = 0
      TSUM = 0.0
      TSUMSQ = 0.0
      WRITE(6,700)
      DO 400 K = 1,N
      RNN = NN
      SUMX = 0.0
      SUMY = 0.0
      SUMXY = 0.0
      SUMXSQ = 0.0
      SUMYSQ = 0.0
      DO 300 J = 1,NN
      X = 0.0
      IF(J.GT.NN/2) X = 1.0
      TEMP = RMS(IST)
      Y = ALPHA + BETA0*X + TEMP
      SUMY = SUMY + Y
      SUMX = SUMX + X
      SUMXY = SUMXY + X*Y
      SUMXSQ = SUMXSQ + X*X
      SUMYSQ = SUMYSQ + Y*Y
  300 CONTINUE
      XBAR = SUMX/RNN
      YBAR = SUMY/RNN
      BETAC = (RNN*SUMXY - SUMX*SUMY)/(RNN*SUMXSQ - SUMX*SUMX)
      SSE = SUMYSQ - RNN*YBAR*YBAR - BETAC/RNN*
     1 (RNN*SUMXY - SUMX*SUMY)
      SVAR = SSE/(RNN - 2.0)
  700 FORMAT(1H148X12HSAMPLE SLOPE 10X15HSAMPLE VARIANCE )
      XINT = TALPHA(I)*SQRT(SVAR/(SUMXSQ - RNN*XBAR*XBAR))
      IF(BETA0.GE.BETAC - XINT.AND.BETA0.LE.BETAC + XINT)
     1 INTEST = INTEST + 1
      WRITE(6,500) K,BETAC,SVAR
      TSUM = TSUM + BETAC
      TSUMSQ = TSUMSQ + BETAC*BETAC
  500 FORMAT(1H 42X13,4XF8.3,15XF8.3)
  400 CONTINUE
      TSLOPE = TSUM/FLOAT(N)
      TVAR = TSUMSQ/FLOAT(N) - TSLOPE*TSLOPE
      WRITE(6,600) INTEST,TSLOPE,TVAR
  600 FORMAT(1H035X13,10XF7.3,10XF7.3)
  900 CONTINUE
      STOP
      END
```


http

	SAMPLE SLOPE	SAMPLE VARIANCE
1	1,369	0,619
2	1,939	1.516
3	0,737	1.293
4	0,791	0.482
5	0,760	2,002
6	0,141	1,406
7	1,105	0,780
8	1,433	2,123
9	1,719	1.861
10	0,349	0.751
11	2,214	1,400
12	1,558	0,266
13	0,134	0,508
14	1,106	1,171
15	0,726	0,617
16	1,049	0,747
17	1,579	0,283
18	1,005	0,573
19	0,385	0,640
20	1,185	0,561
21	0.813	0,720
22	0,012	1,131
23	-0,011	0,653
24	1.215	1,119
25	0,637	1,085
26	0,897	0,626
27	0,996	0,808
28	2,222	1,620
29	0,708	0,664
30	1,037	1,429
31	1,275	0,785
32	0,768	0,690
33	1,193	0,563
34	0,570	1,033
35	0,985	0,838
36	2,386	0,528
37	1,063	1,142
38	1,621	0,807
39	1,724	1,262
40	3,298	2,234
41	1,485	0,872
42	1,755	1,431
43	1,098	1,700
44	1,023	0,586
45	0,283	0,676
46	-0,417	1,045
47	0,989	1,834
48	0,147	1,089
49	1,839	0,790
50	-0,056	0,026
51	0,723	1,291
52	0,962	1,474
53	0,679	0,756
54	0.657	0.649

http

55	0,182	1,126
56	1,180	0,913
57	2,410	1,161
58	0,551	0,573
59	0,610	2,965
60	1,297	1,625
61	1,904	0,603
62	2,158	0,882
63	1,349	1,072
64	1,293	0,611
65	0,611	0,843
66	0,886	0,230
67	0,300	0,601
68	0,680	1,451
69	1,776	1,808
70	0,151	1,010
71	1,209	1,318
72	2,115	1,340
73	-0,741	2,350
74	1,422	0,765
75	0,783	0,964
76	1,265	0,315
77	1,135	0,282
78	0,477	1,572
79	-0,825	1,017
80	0,559	1,184
81	1,839	0,490
82	0,679	0,540
83	1,522	1,746
84	0,460	0,752
85	0,248	0,464
86	1,243	0,761
87	0,427	0,767
88	1,649	2,927
89	-0,653	0,864
90	1,337	1,253
91	0,897	0,466
92	1,128	1,779
93	2,043	0,860
94	-0,139	1,309
95	0,542	0,675
96	0,554	0,462
97	0,534	0,795
98	2,164	0,928
99	1,160	0,534
100	0,761	1,259
95	0,990	0,509

http

	SAMPLE SLOPE	SAMPLE VARIANCE
1	1,806	0.934
2	1,321	1,069
3	0,667	0,657
4	0,592	1,010
5	0,808	0,791
6	0,534	1,155
7	1,266	0,627
8	1,302	1,009
9	0,946	0,924
10	0,862	1,704
11	1,189	1,180
12	0,849	0,806
13	0,997	1,019
14	0,855	1,258
15	0.394	1,209
16	1,108	1,190
17	0,839	1,161
18	1,072	0,773
19	0,664	1,134
20	1,213	0,869
21	0,314	1,254
22	0,830	0,987
23	0,738	1,045
24	0,901	0,878
25	1,257	0,732
26	0,869	0,731
27	1,315	1,008
28	0,790	0,714
29	0,879	0,862
30	1,008	1,398
31	1,347	1,205
32	1,740	0,630
33	0,561	0,987
34	1,296	1,532
35	0,921	1,100
36	0,899	1,160
37	0,940	1,011
38	1,240	0,984
39	1,223	0,781
40	0,920	1,604
41	0,794	1,170
42	0,429	0,714
43	1,259	0,787
44	0,208	1,177
45	0,466	0,719
46	1,042	1,113
47	0,715	0,977
48	0,912	0,994
49	1,182	0,880
50	1,295	0,837
51	1,116	0,903
52	1,359	0,832
53	1,354	0,946
54	0,998	0,915

http

55	1,335	0,765
56	0,700	1,010
57	0,930	1,185
58	0,816	0,923
59	0,884	1,266
60	1,489	1,012
61	0,932	0,936
62	1,187	1,008
63	1,193	1,335
64	1,152	0,825
65	1,052	1,089
66	0,305	0,956
67	0,167	1,124
68	1,644	1,496
69	1,398	1,469
70	0,449	1,554
71	0,856	0,822
72	0,879	0,817
73	0,579	1,308
74	0,617	1,094
75	0,811	1,246
76	0,883	0,772
77	0,606	0,970
78	0,728	0,802
79	1,142	0,719
80	0,450	0,430
81	0,381	0,798
82	1,156	0,949
83	1,021	0,790
84	0,825	0,998
85	0,829	0,958
86	0,912	1,115
87	1,411	1,101
88	1,318	0,990
89	1,503	0,535
90	0,765	0,847
91	1,572	0,811
92	1,272	0,752
93	1,095	0,988
94	0,838	0,857
95	1,156	0,688
96	0,270	0,849
97	1,055	0,719
98	0,867	0,902
99	1,218	0,951
100	1,432	1,060
88	0,965	0,116

http

```
C               EXERCISE 7-4
      DIMENSION BETA(3)
      DATA NN,ALPHA,BETA,BETEST/10,5.0,0.0,0.0,1.0,3.0,0.0,0/
      DATA IST/8675431667/
      RNN = NN
      XP = 1.5
      DO 1000 K = 1,3
      E0 = ALPHA + BETA(K)*XP
      SUMX = 0.0
      SUMY = 0.0
      SUMXY = 0.0
      SUMXSQ = 0.0
      SUMYSQ = 0.0
      DO 300 J = 1,NN
      X = 1.0
      IF(J.GT.NN/2) X = 2.0
      Y = ALPHA + BETA(K)*X + RMS(IST)
      SUMY = SUMY + Y
      SUMX = SUMX + X
      SUMXY = SUMXY + X*Y
      SUMXSQ = SUMXSQ + X*X
      SUMYSQ = SUMYSQ + Y*Y
  300 CONTINUE
      XBAR = SUMX/RNN
      YBAR = SUMY/RNN
      BETAC = (RNN*SUMXY - SUMX*SUMY)/(RNN*SUMXSQ - SUMX*SUMX)
      ALPHAC = YBAR - BETAC*XBAR
      SSE = SUMYSQ - RNN*YBAR*YBAR - BETAC/RNN*
     1 (RNN*SUMXY - SUMX*SUMY)
      SVAR = SSE/(RNN - 2.0)
      YHAT = ALPHAC + BETAC*XP
      T = (YHAT - E0)/SQRT((1.0/RNN + (XP - XBAR)*(XP - XBAR)
     1 /(SUMXSQ - RNN*XBAR*XBAR))/SVAR)
      WRITE(6,600) T
  600 FORMAT(1H055X13HA T-VALUE IS F7.3)
 1000 CONTINUE
      STOP
      END
```

```
A T-VALUE IS  -0.324

A T-VALUE IS  -0.579

A T-VALUE IS  -2.082
```

```
C               EXERCISE 8-1
      DIMENSION YBAR(2),RMU1(2),BETA(2)
      DIMENSION SUMY(2),SUMYSQ(2)
      DATA NN,BETA/50,1.0,1.5/
      DATA IST/8675431667/
      DATA RMU1,BETA/1.0,2.0,2.0,0.0,0.0/
      RNN = NN
      DO 1000 K = 1,2
      DO 900 I = 1,2
      SUMY(I) = 0.0
      SUMERR = 0.0
      SUMYSQ(I) = 0.0
      WRITE(6,700)
      X = FLOAT(I-1)
      DO 300 J = 1,NN
      TEMP = RMS(IST)
      Y = RMU1(K) + X*BETA(K) + TEMP
      SUMY(I) = SUMY(I) + Y
      SUMERR = SUMERR + TEMP
      WRITE(6,800) X,Y
      SUMYSQ(I) = SUMYSQ(I) + Y*Y
  300 CONTINUE
      YBAR(I) = SUMY(I)/RNN
  600 FORMAT(1H0,10X,5(10XF10.3)//)
  700 FORMAT(1H1,52X,27H50 DATA POINTS AS X-Y PAIRS)
  800 FORMAT(1H ,55X,1H(,F4.2,1H,,F5.2,1H))
  900 CONTINUE
      BETAHT = YBAR(2) - YBAR(1)
      EPSAVE = SUMERR/FLOAT(2*NN)
      SEPSSQ = (SUMYSQ(1) + SUMYSQ(2) - RNN*(YBAR(1) +
     1 YBAR(2)))/(RNN - 2.0)
      WRITE(6,600) YBAR(1),BETAHT,EPSAVE,SEPSSQ
 1000 CONTINUE
      STOP
      END
```

```
50 DATA POINTS AS X-Y PAIRS
(0,  , 0,36)
(0,  , 0,90)
(0,  ,-0,27)
(0,  , 0,91)
(0,  , 1,53)
(0,  , 1,04)
(0,  , 2,42)
(0,  , 0,57)
(0,  , 1,18)
(0,  , 0,05)
(0,  , 0,47)
(0,  , 0,45)
(0,  ,-1,97)
(0,  , 1,76)
(0.  , 1,21)
(0,  ,-0,10)
(0,  , 1,14)
(0,  , 1,98)
(0,  , 2,53)
(0,  , 1,06)
(0,  , 0,07)
(0,  , 2,30)
(0,  , 0,08)
(0,  ,-0,33)
(0,  , 0,64)
(0,  , 0,06)
(0,  , 0,36)
(0,  ,-0,93)
(0,  ,-0,36)
(0,  , 2,31)
(0,  , 0,79)
(0,  , 1,58)
(0,  , 2,42)
(0,  , 0,84)
(0,  , 0,34)
(0,  , 0,89)
(0,  , 1,28)
(0,  , 1,21)
(0,  , 1,46)
(0,  , 0,08)
(0,  ,-1,75)
(0,  , 3,03)
(0,  , 2,42)
(0,  , 1,29)
(0,  , 0,95)
(0,  , 1,52)
(0,  , 0,06)
(0,  , 0,94)
(0,  , 1,90)
(0,  , 0,31)
```


http

50 DATA POINTS AS X-Y PAIRS

(1,00, 4,79)
(1,00, 2,34)
(1,00, 3,82)
(1,00, 3,38)
(1,00, 3,36)
(1,00, 2,59)
(1,00, 4,27)
(1,00, 3,69)
(1,00, 0,58)
(1,00, 2,27)
(1,00, 1,72)
(1,00, 3,66)
(1,00, 1,60)
(1,00, 2,65)
(1,00, 4,28)
(1,00, 3,50)
(1,00, 2,96)
(1,00, 2,86)
(1,00, 2,36)
(1,00, 2,77)
(1,00, 5,85)
(1,00, 1,51)
(1,00, 2,02)
(1,00, 1,44)
(1,00, 2,14)
(1,00, 3,30)
(1,00, 2,37)
(1,00, 3,00)
(1,00, 2,06)
(1,00, 4,40)
(1,00, 2,04)
(1,00, 3,12)
(1,00, 0,22)
(1,00, 4,61)
(1,00, 3,11)
(1,00, 3,18)
(1,00, 3,57)
(1,00, 4,97)
(1,00, 2,77)
(1,00, 2,20)
(1,00, 3,71)
(1,00, 3,75)
(1,00, 3,02)
(1,00, 3,26)
(1,00, 2,79)
(1,00, 1,96)
(1,00, 2,76)
(1,00, 4,57)
(1,00, 2,36)
(1,00, 1,63)

0,860	2,083	-0,029	8,201

```
50 DATA POINTS AS X-Y PAIRS
    (0.   , 0.32)
    (0.   , 0.17)
    (0.   , 2.02)
    (0.   , 3.18)
    (0.   , 1.19)
    (0.   , 0.88)
    (0.   , 3.52)
    (0.   , 3.40)
    (0.   , 2.10)
    (0.   , 3.05)
    (0.   , 0.60)
    (0.   , 1.47)
    (0.   , 1.37)
    (0.   , 2.24)
    (0.   , 2.25)
    (0.   , 2.52)
    (0.   , 2.25)
    (0.   , 1.99)
    (0.   , 2.02)
    (0.   , 1.95)
    (0.   , 3.17)
    (0.   , 2.75)
    (0.   , 1.47)
    (0.   , 2.56)
    (0.   , 1.64)
    (0.   , 2.67)
    (0.   , 0.95)
    (0.   , 1.21)
    (0.   , 1.28)
    (0.   , 1.15)
    (0.   , 1.50)
    (0.   , 2.91)
    (0.   , 2.43)
    (0.   , 1.53)
    (0.   , 2.62)
    (0.   , 0.60)
    (0.   , 2.96)
    (0.   , 2.30)
    (0.   , 4.22)
    (0.   , 1.46)
    (0.   , 1.39)
    (0.   , 2.51)
    (0.   , 2.44)
    (0.   , 1.35)
    (0.   , 2.97)
    (0.   , 1.96)
    (0.   , 2.14)
    (0.   , 0.53)
    (0.   , 1.80)
    (0.   , 2.85)
```


http

50 DATA POINTS AS X-Y PAIRS

(1,00, 1,59)
(1,00, 2,07)
(1,00, 2,60)
(1,00, 1,59)
(1,00, 0,03)
(1,00, 2,04)
(1,00, 0,89)
(1,00, 2,27)
(1,00, 2,20)
(1,00, 0,71)
(1,00, 1,32)
(1,00, 1,42)
(1,00, 2,76)
(1,00, 1,81)
(1,00, 2,32)
(1,00, 2,97)
(1,00, 2,07)
(1,00, 2,92)
(1,00, 2,06)
(1,00, 2,51)
(1,00, 3,84)
(1,00, 2,05)
(1,00, 1,93)
(1,00, 2,34)
(1,00, 1,69)
(1,00, 3,27)
(1,00, 1,76)
(1,00, 2,68)
(1,00, 1,77)
(1,00, 2,39)
(1,00, 0,76)
(1,00, 3,02)
(1,00, 2,33)
(1,00, 1,48)
(1,00, 3,25)
(1,00, 1,41)
(1,00, 1,74)
(1,00, 2,19)
(1,00, 1,04)
(1,00, 1,38)
(1,00, 1,37)
(1,00, 1,54)
(1,00, 1,28)
(1,00, 0,98)
(1,00, 1,80)
(1,00, 0,96)
(1,00, 0,82)
(1,00, 1,94)
(1,00, 3,21)
(1,00, 0,97)

1,996	-0,089	-0,046	5,319

http

36 W1000

http://www.brainless.org/Altair/